

Long-Memory and Level Shifts in the Volatility of Stock Market Return Indices*

Pierre Perron[†]

Zhongjun Qu[‡]

Boston University

Boston University

February 4, 2008. This version: September 17, 2008

Abstract

There has been interest in the possibility of confusing long memory and structural changes in level and studies showed that when a short memory process is contaminated by level shifts the estimate of the fractional differencing parameter is biased upward and the autocovariances decay slowly. We analyze the properties of the autocorrelation function, the periodogram and the log periodogram estimate of the memory parameter when the level shift component is a mixture model. Our results explain many findings reported and uncover new features. We confront our theoretical predictions using log-squared returns as a proxy for the volatility of some assets returns, including daily S&P 500 returns over the period 1928-2002. The estimates follow patterns that would obtain if the process was short-memory with level shifts instead of fractionally integrated. A simple test is also proposed, which reinforces this conclusion.

JEL Classification Number: C22.

Keywords: structural change, jumps, long memory processes, fractional integration, frequency domain estimates.

*This is a revised version of parts of a working paper entitled “An Analytical Evaluation of the Log-periodogram Estimate in the Presence of Level Shifts and its Implications for Stock Returns Volatility”. We thank the participants of the Harvard/MIT Econometrics Seminar, Serena Ng, a referee and, especially, Yohei Yamamoto for useful comments as well as Shinsuke Ikeda for research assistance. Pierre Perron acknowledges financial support from the National Science Foundation under Grant SES-0649350.

[†]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (perron@bu.edu).

[‡]Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (qu@bu.edu).

1 Introduction

There has long been an interest in long memory models, especially with applications to financial times series. For reviews of the literature, see Robinson (1994), Beran (1994) and Baillie (1996). Of particular interest in econometrics is the fractionally integrated model (Granger, 1981, Granger and Joyeux, 1980, and Hosking, 1981) whose difference of order d is a short-memory process often modelled as an *ARMA* process for which the autocorrelations decays exponentially. The parameter d can be non-integer valued and when $0 < d < 1$, the autocorrelations decays very slowly, a characteristic of long memory processes. Various methods have been proposed to estimate the long memory parameter d . One often used is a semiparametric estimator in the frequency domain which does not require a distributional assumption on the process generating the difference of order d of the series. A popular method is the log periodogram regression proposed by Geweke and Porter-Hudak (1983), whose large sample distribution was analyzed by, among others, Robinson (1995) and Hurvich, Deo and Brodsky (1998) as well as Phillips (2007) who covers the unit root case $d = 1$. Applications to macroeconomics, international trade and finance are numerous. For example, Ding et al. (1993) argue that stock returns volatility is well described by a long memory process.

Recently, there has been an upsurge of interest on the possibility of confusing long memory and structural changes in level. The idea extends that expositied in Perron (1989, 1990) who showed that structural changes and unit roots ($d = 1$) are easily confused in the sense that, with a stationary process contaminated by structural changes, the estimate of the sum of the autoregressive coefficients is biased towards 1 and that tests of the null hypothesis of a unit root are biased towards non-rejection. This phenomenon has been shown to apply in the long memory context as well. When a stationary short memory process is contaminated by structural changes in level the estimate of d is biased away from 0 and the autocovariance function exhibits a slow rate of decay. Relevant references on this issue include Diebold and Inoue (2001), Engle and Smith (1999), Gouriéroux and Jasiak (2001), Granger and Ding (1996), Granger and Hyung (2004), Lobato and Savin (1998), Mikosch and Stărică (2004a), Parke (1999) and Teverosovky and Taqqu (1997). While some papers contain theoretical results related to the fact that the variance and autocorrelations have similar properties under structural change and long memory, most of the evidence was obtained through simulations and no theoretical results are available pertaining to the distribution of the estimate of the long memory parameter d in the presence of a short memory process with level shifts.

The specification adopted in this paper is one for which the series of interest is the sum

of a short memory process and a jump or level shift component. For the latter, we specify the commonly used simple mixture model such that the component is the cumulative sum of a process which is 0 with some probability $(1 - p/T)$ and is some random variable with probability p/T . Level shifts then occur with some probability p/T that we make dependent on the sample size to obtain non-degenerate limiting results. The underlying idea is to have infrequent changes that are more akin to structural changes rather than a large number of changes which would make the level shift component basically an integrated process ($d = 1$). By scaling the probability of a level shift by T , this aim is achieved in large samples (of course, this specification has no effect in finite samples). This allows us to use a Functional Central Limit Theorem recently obtained by Georgiev (2002) and Leipus and Viano (2003).

We present theoretical results about the limit distributions of the autocorrelation function and the periodogram. We shall also use theoretical results derived in Perron and Qu (2007) about the log periodogram estimate of d . These will allow us to explain many of the findings reported in the literature mentioned above. Moreover, it also allows us to uncover new features and gain better insight about the properties of the various estimates. In particular, we show that the reliance on using the familiar rule of thumb $T^{1/2}$ for the number of frequencies used to estimate the regression, e.g. Diebold and Inoue (2001), allows only a very narrow picture of the problem. We explain how the limit distribution of the log periodogram estimate is highly dependent on the number of frequencies used, a feature that is different from the case where the true underlying process is a pure fractionally integrated model. Hence, this can be helpful to distinguish structural change from long memory.

Our theoretical results have important practical implications that can be confronted with the data. An area where such a concern is important pertains to the behavior of stock return volatility. Several papers have reported that transformations of returns, r_t , of the form $|r_t|^\theta$ for some $\theta > 0$ have times series properties that resemble those of a long-memory process (see, e.g., Ding, Engle and Granger, 1993, Granger and Ding, 1996 and Lobato and Savin, 1998). Similar evidence applies to log absolute returns (see Stărică and Granger, 2005). It has also been documented using realized volatility constructed from high frequency data for various assets (e.g., Andersen and Bollerslev, 1997, Bollerslev and Wright, 2000 and Andersen, Bollerslev, Diebold and Labys, 2003, among many others).

More recently, attempts have been made to distinguish between the stationary noise plus level shift and the long-memory models; see, in particular, Granger and Hyung (2004). They document the fact that, when breaks determined via some pre tests are accounted for, the evidence for long memory is weaker. This evidence is, however, inconclusive since

structural change tests are severally biased in the presence of long memory and the log periodogram estimate is biased downward when sample-selected breaks are introduced. This is an overfitting problem that Granger and Hyung (2004, p. 416) clearly recognized. Stărică and Granger (2005) present evidence that log-absolute returns on the S&P 500 index is an *i.i.d.* series affected by occasional shift in the unconditional variance and show that such a specification has a better forecasting performance than the more traditional GARCH(1,1) model. Mikosch and Stărică (2004b) consider the autocorrelation function of the absolute returns on the S&P 500 index for the period 1953-1977. They document the fact that for the full period, it resembles that of a long memory process. But, interestingly, if one omits the last four years of data, the autocorrelation function is then very different and looks like one associated with a short memory process. They explain this finding by arguing that the volatility of S&P 500 returns has increased over the period 1973-1977.

Our results allow a way of discriminating between the two models based on the autocorrelation function and the path of the log periodogram estimates \hat{d} as m varies. We shall first illustrate the implications using the same data set as in Granger and Hyung (2004). It consists of 19,868 daily observations for the S&P 500 returns over the period January 4, 1928 to October 30, 2002. It was kindly provided by William Schwert. The source of the data for the period January 4, 1928 through July 2, 1962 is Schwert (1990). From July 3, 1962 it is from the CRSP daily returns file. We use log returns ($r_t = 100[\ln(P_t) - \ln(P_{t-1})]$ with P_t the S&P 500 stock price index) whose series is depicted in Figure 1. As a proxy for volatility, we use log-squared returns following Stărică and Granger (2005). To account for the presence of zero returns, we use the following measure $\ln(r_t^2 + 0.001)$. The conclusions remain the same if we eliminate returns that are less than 0.000001, which involves 379 observations. The results are also similar using absolute or square root returns. The sample path of the log periodogram estimates follows a pattern that would obtain if the underlying process is one of short-memory with level shifts. We also show that the autocorrelation function of a short-memory process with level shifts has a special structure, in particular for large lags its shape (in expected value) depends only on the sample size in contrast to that of a long-memory process, whose autocorrelation function has a shape dictated by the underlying process. When analyzed in its entirety, the behavior of the sample autocorrelation function of stock return volatility also favors a short-memory process with level shifts. In Section 4, we consider sensitivity analyses to various sub-samples and returns series on other indices, and obtain similar conclusions. We finally propose a simple test of the null hypothesis of a long memory which has power against a level shift process. It is based on the difference in

the estimates \hat{d} constructed using different numbers of frequencies. The test applied to stock return volatility shows a rejection of the null hypothesis of a long memory process.

The structure of the paper is as follows. Section 2 describes the data generating process used throughout and the Functional Central Limit Theorem for the level shift process. Section 3 considers the limiting distributions of the autocovariance function and of the periodogram with simulations showing that they provide good approximations. We also summarize the relevant results from Perron and Qu (2007) about the properties of the log-periodogram estimate of d . In all cases, we consider how well the theoretical predictions fit the data when using the series of log squared returns on the S&P 500 index. Section 4 documents via simulations calibrated to estimates obtained from S&P 500 returns over the period 1980.1-2005.12, that a simple level shift model can easily explain the theoretical and empirical features documented. Section 5 analyzes via simulations whether standard long-memory models, with noise and/or with level shifts, can explain the key documented features in the data. Section 6 proposes simple tests for the null hypothesis of long memory, which have power against an alternative of a short-memory process with level shifts. Section 7 offers brief conclusions and a mathematical appendix some technical derivations.

2 The data generating process with mean shifts

The data generating process adopted is quite simple, yet rich enough to provide theoretical explanations for many of the simulation results about the effect of level shifts on long memory parameter estimates. It is a mixture of a short memory process and a component determined by shifts occurring according to a Bernoulli process. More specifically, DGP-1 is

$$x_t = c + v_t + u_{T,t}, \quad u_{T,t} = \sum_{j=1}^t \delta_{T,j}, \quad \delta_{T,t} = \pi_{T,t} \eta_t. \quad (1)$$

Here c is a constant and v_t is a short memory process defined by $v_t = C(L)e_t$ with $e_t \sim i.i.d.$ $(0, \sigma_e^2)$ and $E|e_t|^r < \infty$ for some $r > 2$. The polynomial $C(L)$ satisfies $C(L) = \sum_{i=0}^{\infty} c_i L^i$, $\sum_{i=0}^{\infty} i|c_i| < \infty$ and $C(1) \neq 0$. For the level shift component, $\eta_t \sim i.i.d.$ $(0, \sigma_\eta^2)$ and $\pi_{t,T}$ is a Bernoulli variable that takes value 1 with probability p/T , i.e. $\pi_{T,t} \sim i.i.d.$ $B(p/T, 1)$. We also assume that the components $\pi_{T,t}$, η_t , and v_t are mutually independent.

Remark 1 *Note that p is independent of the sample size T . This specification is needed to model structural changes in mean, i.e., relatively infrequent events that affect the properties of the series in a permanent fashion. If p/T converges to some value in $(0, 1)$, the model is best construed as depicting a standard unit root process.*

Remark 2 *We make no claim that model (1) is the true or the best description of the data generating process for any of the series analyzed. We use this very simple model on purpose since the goal is to distinguish between level shifts plus short-memory noise and long-memory and what is needed are the essential elements. All models are approximations and the usefulness of a particular specification is to be assessed on how well the theoretical predictions fits the data. If level shifts are an important element, we should expect the estimates to behave according to our theoretical predictions, at least to a first approximation. If the model is not a useful approximation, our theoretical predictions will not hold empirically.*

A crucial ingredient that will be used throughout the paper is a Functional Central Limit Theorem for the cumulative level shifts process $u_{T,t}$. This has been considered by Georgiev (2002) and Leipus and Viano (2003). The results relevant to our analysis are stated in the following Lemma where “ \Rightarrow ” denotes weak convergence under the Skorohod topology.

Lemma 1 *(Georgiev, 2002; Leipus and Viano, 2003) Consider DGP-1 with $0 < p < \infty$, then $u_T(s) = \sum_{t=1}^{\lfloor Ts \rfloor} \delta_{T,t} \Rightarrow J(s)$ where $J(s) = \sum_{j=0}^{N(s)} \eta_j$ with $N(s)$ a Poisson process with jump intensity p which is independent of η_j for all j .*

Remark 3 *The limiting distribution $J(s)$ depends on the exact distribution of η_t . Below, to obtain quantitative results to assess important features of the distributions and their adequacy as approximations to the finite sample distributions, we shall specify a normal distribution.*

Since we shall make frequent comparisons with long memory processes, it is useful to make precise the properties we shall refer to. Here, we will use the following two definitions of a long memory process. Let $\{x_t\}_{t=1}^T$ be a stationary time series with spectral density function $f_x(w)$ at frequency w , then x_t is said to have long memory if

$$f_x(w) = g(w)w^{-2d} \text{ as } w \rightarrow 0 \quad (2)$$

with $g(w)$ a slowly varying function as $w \rightarrow 0$ (i.e., for any real t , $g(tw)/g(w) \rightarrow 1$ as $w \rightarrow 0$). When $d > 0$, this implies that the spectral density function increases for frequencies that get close to zero. The rate of divergence to infinity depends on the parameter d . Under some general conditions, this low-frequency definition is equivalent to the following long-lag autocorrelation definition (Beran, 1994). Let $\gamma_x(\tau)$ be the autocorrelation function of x_t . If $\gamma_x(\tau) = c(\tau)\tau^{2d-1}$ as $\tau \rightarrow \infty$, with $c(\tau)$ a slowly varying function as $\tau \rightarrow \infty$, the process is said to have long memory. For $0 < d < 1/2$, this implies that the autocorrelations decreases to zero at a slow hyperbolic rate which depends on the parameter d , in contrast to the fast geometric rate of decay that applies to a short-memory process.

3 The limit distributions of the various statistics

In this section, we examine the properties of some statistics in both the time and frequency domains under the DGP-1. We consider the autocorrelation function, the periodogram, and summarize relevant results about the limit distribution of the log-periodogram estimate of the memory parameter derived in Perron and Qu (2007). We consider how well the theoretical predictions fit the data when using the series of log squared returns on the S&P 500 index.

3.1 The sample autocovariance function

With an unknown mean, the sample covariance at lag h is defined by $\hat{R}(h) = T^{-1} \sum_{t=1}^{T-h} (x_t - \bar{x})(x_{t+h} - \bar{x})$ with $\bar{x} = T^{-1} \sum_{t=1}^T x_t$. We shall study the properties of $\hat{R}(h)$ as $T \rightarrow \infty$ under two scenarios for the relation between h and T : a) with $h/T \rightarrow 0$ as $T \rightarrow \infty$ (fixed- h asymptotic), and b) with $h/T \rightarrow \kappa$ as $T \rightarrow \infty$ (large- h asymptotic). The result for case (a) is stated in the following Proposition proved in the appendix.

Proposition 1 *Under DGP-1, if $h/T \rightarrow 0$ as $T \rightarrow \infty$, $\hat{R}(h) \Rightarrow R_v(h) + \int_0^1 (J(s) - \bar{J})^2 ds$, where $\bar{J} = \int_0^1 J(s) ds$ and with $R_v(h)$ the autocovariance function of v_t .*

The limiting distribution has two components. The first is the standard autocovariance function of the short memory process. The second corresponds to the cumulative level shift process and is a positive random variable that is independent of the lag h . Hence, for h small the former will dominate but since it eventually decreases at an exponential rate, the second component will dominate for h large. Accordingly, the limit will exhibit a very slow rate of decrease which is a characteristic of a long memory process. Technically, the limiting value does not decrease to zero as h increases contrary to a stationary process. But recall that given the condition $h/T \rightarrow 0$ as $T \rightarrow \infty$, the limiting distribution is not tailored to provide a good approximation for large values of h . This can be achieved by considering the limit assuming $h/T \rightarrow \kappa$ as $T \rightarrow \infty$, which is stated in the following Proposition.

Proposition 2 *Under DGP-1, if $h/T \rightarrow \kappa$ as $T \rightarrow \infty$, then for $0 < \kappa < 1$, $\hat{R}([T\kappa]) \Rightarrow \int_0^{1-\kappa} (J(s) - \bar{J})(J(s + \kappa) - \bar{J}) ds \equiv R(\kappa)$ and $E(R(\kappa)) = (p\sigma_\eta^2/6)[(1 - \kappa)^3 - 3(1 - \kappa^2) + (1 - \kappa^3) + 2(1 - \kappa)]$.*

Note that the short memory component is no longer present, since the autocovariance of a short memory process decays exponentially. Hence, for large h the autocovariance function is influenced solely by the Poisson process. The functional described in Proposition 2 is

strictly decreasing as κ increases, though very slowly. Combining Propositions 1 and 2, we have the following result for the autocorrelation function for $0 < \kappa < 1$:

$$\rho([T\kappa]) \Rightarrow \int_0^{1-\kappa} (J(s) - \bar{J}) (J(s + \kappa) - \bar{J}) ds / [\int_0^1 (J(s) - \bar{J})^2 ds + R_v(0)].$$

To assess the adequacy of the asymptotic approximations, Figure 2 presents the results of a simple simulation experiment. The short-memory component is an $AR(1)$ process of the form $v_t = \rho v_{t-1} + e_t$ with $e_t \sim i.i.d. N(0, 1)$ and $\rho = 0.5, 0.9$. The level shift component is specified by (1) with $p = 5$ and $\eta_t \sim i.i.d. N(0, 1)$. The sample size is $T = 500$, the number of replications is 10,000 and the first 150 autocorrelations are plotted. We present the median of the exact values of the autocorrelations and the fixed- h and large- h approximations. The results show interesting features. First, as documented elsewhere, the finite sample autocorrelations decrease very slowly, in a way similar to a long-memory process. Consider now the adequacy of the two asymptotic approximations. For small values of h , the fixed- h asymptotic provides a good approximation while the large- h asymptotic is not satisfactory, and vice versa for large values of h . The values of h for which one or the other approximation is good depend on the correlation in the short-memory component. When $\rho = 0.5$, the fixed- h asymptotic is good until lag 4 and the large- h asymptotic is good for h larger than 5. When $\rho = 0.9$, the fixed- h is good until roughly lag 20, while the large- h asymptotic is good for h larger than 40. Hence, both asymptotic approximations are complementary.

Even though a short-memory process with level shifts has an autocorrelation structure that resembles that of a stationary long memory process for a large number of lags, the two are not observationally equivalent. Since the feature that characterizes the autocorrelation function of a long memory process is its behavior for distant lags, it is appropriate to use the result of Proposition 2 which yields a good approximation in this case. Consider then the crude approximation of the autocorrelation function given by $E(R(\kappa))/E(R(0))$. For a short-memory process with level shifts, an approximation to the autocorrelations (for large lags) as a function of κ is $f(\kappa)/(1+a)$ where $f(\kappa) = [(1-\kappa)^3 - 3(1-\kappa^2) + (1-\kappa^3) + 2(1-\kappa)]$ and $a = 6R_v(0)/(p\sigma_\eta^2)$. This is an important result, which can have useful testable implications. Note first that $f(\kappa)$ is the approximation that would prevail if the process was a random walk (see Wichern, 1973). Since $a > 0$, the autocorrelation with level shifts is a flattened version that varies between zero when $p = 0$ and approaches $f(\kappa)$ as p and σ_η^2 increases, as expected. The testable implication is that, apart from this scaling, the general shape depends only on T , the sample size. The autocorrelations initially (almost) linearly decrease (as depicted in the simulated values in Figure 2). It then crosses the zero axis when $\kappa = 1 - \sqrt{1/2} \approx 0.293$,

reaches a minimum at $\kappa = 1 - \sqrt{1/6} \approx 0.592$ and increases up to 0 when $\kappa = 1$. The sample autocorrelations of a long memory process is different and crosses the zero axis earlier at a location depending on the memory parameter (see Section 5).

We now consider the autocorrelation function of the log squared returns of the S&P 500 index as a proxy for its volatility. The top panel of Figure 3 presents the autocorrelation function up to lag 2,500. It is this kind of graphical representation that has lead many researchers to conclude that a long memory process is a relevant contender to explain stock return volatility. The decline appears to be slow and the values are above zero even for lags as far as 2,000. But with 19,868 daily observations, restricting the analysis to the first 2,500 lags does not allow us to depict the important features discussed above. To that effect, the bottom panel of Figure 3 presents the autocorrelation functions for all lags. The shape is almost exactly as predicted by our theory if the underlying process is one of short-memory with level shifts. It reaches a zero value at roughly lag 5,000 (approximately $.25T$), reaches a minimal value at lag 9,000 and increases back up to 0 at the most distant lag.

But we can confront the data with additional testable implications, in particular the fact that the autocorrelation function has the same shape irrespective of the size of the sample. We consider the autocorrelation functions for various sub-samples that all start in 1928 (with the first observation available). To ease presentation, we smoothed the autocorrelation functions using a non-parametric kernel smoothing method with a normal kernel and the bandwidth set to $T^{-1/3}$. The results are presented in Figure 4 for samples of sizes 3, 6, 9 and 18 thousands daily observations. Again, they fit the theory well. All functions initially decrease and cross the zero axis at roughly $.3T$, reach a minimum around $0.6T$ and go back to zero for the most distant lags. This has important implications. For instance, consider the value of the autocorrelation at, say, lag 1,000. With $T = 3,000$ it is negative (roughly -0.1), with $T = 6,000$ it is basically zero and with $T = 9,000$ or $18,000$ it is positive (roughly 0.1). Hence, the strength of the evidence about long memory changes drastically with the size of the sample, being stronger the larger the sample size (the same features hold changing the sample size using a different sampling frequency keeping the same total span of the data).

3.2 The periodogram

We now consider the behavior of the periodogram under DGP-1. Simulation results presented in the literature (e.g., Diebold and Inoue, 2001) have documented the fact that semiparametric methods, such as the log periodogram regression, yield estimates of the memory parameter d significantly above 0 when the DGP is a short memory process with level shifts. We now

provide a theoretical analysis of the underlying components of these estimates, namely the periodogram ordinates (we return to the log periodogram regression in Section 3.3).

The periodogram provides a measure (though imprecise) of the contribution to the total variability of the series from components at different frequencies. For a series x_t and a frequency $w_j = 2\pi j/T$ ($j = 1, \dots, [T/2]$), it is defined by $I_{x,T}(w_j) = (1/2\pi T) |\sum_{t=1}^T x_t \exp(iw_j t)|^2$, where $i = \sqrt{-1}$ and $|\cdot|^2$ stands for the complex conjugate product. The following decomposition using the fact that $x_t = c + v_t + u_{T,t}$ will be useful:

$$\begin{aligned} I_{x,T}(w_j) &= I_{v,T}(w_j) + I_{u,T}(w_j) + 2I_{vu,T}(w_j) \\ &= \frac{1}{2\pi T} \left| \sum_{t=1}^T v_t \exp(iw_j t) \right|^2 + \frac{1}{2\pi T} \left| \sum_{t=1}^T u_{T,t} \exp(iw_j t) \right|^2 + \frac{2}{2\pi T} \sum_{t=1}^T \sum_{s=1}^T v_t u_{T,s} \cos(w_j(t-s)). \end{aligned}$$

Hence, for a particular frequency the contribution to the total variability can be due to three sources: the short memory process, the level shift process and the interaction between the two. Note that given our assumption that v_t and $u_{T,t}$ are independent, the last term has mean zero and has no contribution beyond random variation. We have the following Proposition concerning the order of these three terms.

Proposition 3 *For $j = 1, \dots, [T/2]$, 1) $I_{v,T}(w_j) = (2\pi T)^{-1} |\sum_{t=1}^T v_t \exp(iw_j t)|^2 = O_p(1)$; 2) $\lim_{T \rightarrow \infty} E[(j^2/T)I_{u,T}(w_j)] = p\sigma_\eta^2/4\pi^3$, and the limiting variance is independent of j and bounded; hence, $I_{u,T}(w_j) = O_p(Tj^{-2})$. Also, for a fixed j*

$$T^{-1}I_{u,T}(w_j) \Rightarrow (1/2\pi) \int_0^1 \int_0^1 J(u)J(s) \cos(2\pi j(s-u)) ds du;$$

3) $I_{vu,T}(w_j) = O_p(T^{1/2}j^{-1})$, $\lim_{T \rightarrow \infty} E[(j/T^{1/2})I_{vu,T}(w_j)] = 0$, and, with η_t normally distributed, the limiting variance is $p^2\sigma_\eta^2/4\pi^4$.

Proposition 3 goes a long way towards explaining some of the simulation results in the literature. Consider first the relative magnitude of each term for “small” frequencies in the sense that j satisfies $j = o(T^{1/2})$ as T gets large. In this case, the second term, corresponding to the contribution made by the cumulative level shift component, dominates. As j increases, the rate of decrease of this component is a random variable with mean -2 and finite variance (as we shall see in Section 3.3, this implies that the log-periodogram estimate of the fractional difference parameter should be close to 1 when evaluated using specific numbers of frequencies). Deviations from this rate decrease with increases in either p or σ_η^2 (since the importance of the second term increases with increases in p or σ_η^2). For “large”

frequencies that satisfy $T/j^2 = o(1)$, the second and third components are then $o_p(1)$ and the first component dominates, even though it is itself small.

The results of Proposition 3 are also useful to highlight the differences between a short-memory process with level shifts and a stationary long-memory process. For frequencies that approach zero, and any given fixed T , the periodogram diverges at rate w_j^{-2} , which from (2) is the rate that would apply for a unit root process. Furthermore, as the frequency increases, the periodogram of a short-memory process with level shifts decreases faster than that of a long memory process. Hence, it will be easier to distinguish between the two in the frequency domain rather than the time domain. This is mainly due to the fact that the differences in the autocorrelation functions for distant lags sum up to large differences and imply a different behavior of the spectral density near the origin.

It is important to note the fact that the impact of mean shifts on the periodogram occurs only at frequencies very close to 0. For large frequencies, the first term dominates and $I_x(w_j)$ in large samples has mean $f_v(w_j)$, the spectral density of the short memory component at frequency w_j . To assess how rapidly the impact of mean shifts reduces as the frequency increases, consider DGP-1 with $v_t \sim i.i.d. N(0, 1)$ so that $f_v(w_j) = 1/2\pi$ and $u_{T,t}$ generated from a process with $p = 10$ and $\sigma_\eta^2 = 1$. For $T = 500$, the ratio $E(I_{u,T}(w_j))/E(I_{v,T}(w_j))$ then takes the following values: 20.2 ($j = 1$), 5.04 ($j = 2$), 2.24 ($j = 3$), 1.26 ($j = 4$), 0.81 ($j = 5$) and 0.56 ($j = 6$). Hence, the mean shift component dominates the short memory components only for the first four frequencies up to $w_4 = \pi/63.5$.

To better highlight the relative importance of each component, Figure 5 presents the medians of the three components from a simulation with 10,000 replications. We use DGP-1 with the parameters set to $T = 500$, $\eta_t \sim i.i.d. N(0, 1)$, and $p = 5$. The short memory process is an $AR(1)$ with $i.i.d. N(0, 1)$ errors and autoregressive parameter 0.7. The results show that the jump component is clearly dominant for short frequencies but that beyond a few frequencies the short-memory component dominates (throughout, the cross component has no first-order effect). The jump component has indeed a very important effect on the periodogram but this effect is only in a very narrow band close to frequency zero. In our example, this effect is smaller than the contribution of the short-memory component beyond frequency $\pi/80$ and basically nil beyond frequency $\pi/34$. Taking these results into consideration, we conclude that if we use a local method to estimate the memory parameter of a mean shifting process, the estimate will depend on the number of frequencies used. It will tend to be less affected by mean shifts if we increase the number of frequencies. A plot of the periodogram for the S&P 500 log squared returns is presented in Figure 6. The shape

is exactly what one would expect given the results discussed. Only the first few frequencies are important, but they have an overwhelming dominance compared to higher frequencies. Note that the values continue to be small beyond the maximal frequency $\pi/50$ reported.

3.3 The log periodogram estimate of d

The log periodogram regression estimator of the memory parameter d was proposed by Geweke and Porter-Hudak (1983). It is a semi-parametric method which uses only frequencies near zero to avoid possible misspecification caused by high frequency movements. With $I_{x,T}(w_j)$ the sample periodogram at the j th Fourier frequency $w_j = 2\pi j/T$ ($j = 1, \dots, [T/2]$), the estimate is obtained from the following regression estimated by least-squares

$$\log(I_{x,T}(w_j)) = c - 2d \log(2 \sin(w_j/2)) + \varepsilon_j$$

using observations pertaining to frequencies ranging from $j = 1$ to m . Here m acts as an upper bound on the number of frequencies used. A popular rule of thumb is $m = T^{1/2}$. As a matter of notation, let $a_j = -\log(2 \sin(w_j/2)) + (1/m) \sum_{k=1}^m \log(2 \sin(w_k/2))$ and $S_T = \sum_{k=1}^m a_k^2$. The estimate of d is then $\hat{d} = (1/2S_T) \sum_{j=1}^m a_j \log(I_{x,T}(w_j))$.

We summarize the main features of the limit distributions of \hat{d} derived in Perron and Qu (2007) when the underlying process is the level shift model with short-memory. First, when m is near $T^{1/3}$, \hat{d} will be in a neighborhood of 1 with a standard deviation of about $.79/\sqrt{m}$ (provided $p \geq 5$). When m is roughly between $T^{1/3}$ and $T^{1/2}$, \hat{d} drops to a new level when the stationary component starts to affect the limiting distribution. The magnitude of the drop depends on the relative variance of the stationary and level shift components as well as on the value of p , the frequency of the jumps. As m increases beyond $T^{1/2}$ there is a further gradual decrease in \hat{d} as the short-memory component becomes increasingly more important, relative to the level shift component, in determining the limiting distribution. Note that the rate of increase of m relative to T for which the limit of \hat{d} is one as derived in Perron and Qu (2007) is a lower bound somewhat constrained by the method of proof. What is important is that there is a discontinuity in the asymptotic distribution for small and larger rates of increase of m . In what follows, we adopt $T^{1/3}$ as the bound since it allows large enough values of m for the sample sizes of the series analyzed.

The picture is very different if the underlying model is that of a long-memory process, e.g., a fractionally integrated model. Here, the limiting distribution of the log periodogram estimate \hat{d} is the same regardless of the rate of increase of m relative to the sample size T . Indeed, from Hurvich, Deo and Brodsky (1998), we have, for Gaussian processes, $\sqrt{m}(\hat{d} -$

$d) \rightarrow^d N(0, \pi^2/24)$ if $m = o(T^{4/5})$ and $\log^2(T) = o(m)$. The same result holds for non-Gaussian processes under more stringent conditions on the rate of increase of m (see Deo and Hurvich, 2001). Hence, we can use the path of the estimates \hat{d} obtained for a wide range of values of m to discriminate between the two models.

Consider now the estimates applied to S&P 500 log-squared returns. The log periodogram estimates \hat{d} were computed for all values of m ranging from 10 to $T^{3/4}$. The results are presented in Figure 7. The path of \hat{d} as m varies is almost exactly as predicted by the theoretical results which apply if the true underlying structure is a short-memory process with level shifts. Consider first values of m near $T^{1/3}$ ($m = 27$, indicated by the first vertical solid line from the left). The estimates \hat{d} are clearly highest in this region with a peak at 0.74. This value is between the 5% and 10% quantiles of the asymptotic distribution of Theorem 1 provided $p \geq 5$ (if p is smaller these quantiles would be lower, see Perron and Qu, 2007). Hence, these cannot be viewed as significantly different from 1. Furthermore, the results show that before m takes values near $T^{1/2}$ ($m = 141$, indicated by the second vertical solid line), the estimate drops suddenly. It reaches a value of 0.49. This is close to estimates reported in the literature since the rule of thumb $m = T^{1/2}$ is often used (see, e.g., Granger and Ding, 1996). After $m = T^{1/2}$, the estimates steadily decline. The decrease between $T^{1/2}$ and $T^{2/3}$ ($m = 734$, third vertical solid line) is from 0.49 to 0.36. After $T^{2/3}$, the estimates continue to decline though very slowly. When $m = T^{.8}$ (2,745) we have, from results not reported in the graph, a value less than 0.25.

Can these results be explained by finite sample biases of the log-periodogram estimates that would prevail if the true underlying process was one of long-memory? The answer from what is available in the literature is no. For the standard case where the series is a purely fractionally integrated process, the shape of the bias is basically the same across value of d , it is small if the short-memory component is weakly correlated, and it is indeed important when the short-memory component is strongly serially correlated. But the bias is positive or negative (depending on the second derivative of the spectral density function at frequency zero of the short-memory component) and increasing with m (see Hurvich, Deo, Brodsky, 1998, and especially Figure 1 in Andrews and Guggenberger, 2003). For the so-called long memory stochastic volatility model or perturbed fractional process, i.e., the sum of a fractionally integrated process plus a stationary noise, a more likely scenario given that the proxies used for volatility are noisy measures of the true volatility process, the bias is negative but it is negative *for all* values of m (see, Deo and Hurvich, 2001, Sun and Phillips, 2003, Breidt, Crato and de Lima, 1998, and Hurvich, Moulines and Soulier, 2005). So both

cases cannot explain estimates in the nonstationary region for small m and decreasing in the stationary region for large m . We consider more detailed simulations in Section 5.

It may be argued that our results are specific to the series analyzed and the sample period considered. We assessed the sensitivity of our findings in several directions using: a) the same series with different sub-samples; b) other market indices. The inclusion of the great depression in our sample may be responsible for the main findings as this period may be atypical with large fluctuations in variance. Hence, we considered the following more recent periods: 1957-2002, 1973-2002 and 1990-2002 (the last 3,000 observations). They show again the same results to hold. Indeed, in the most recent period, the results are even more striking as presented in Figure 8. For other market indices, we considered the (log-squared) daily returns on the following indices and sample periods (available from the Wharton Research Data Services): the NASDAQ (1972:12:15 to 2006:12:31; value-weighted returns), the AMEX (1961:07:03 to 2006:12:31; value weighted returns), and the Dow Jones (1957:03:04 to 2002:10:20). The results (available in the working paper version) show that the same patterns hold again, especially for the path of the log-periodogram estimates. Hence, our results are robust to different sample periods and different market indices.

4 Can a simple level shift model explain the empirical features?

To examine whether the features of the sample autocorrelations and the path of the estimates of d as a function of m can be reproduced by the short memory plus level shifts model, we consider a simulation experiment calibrated to empirical estimates. Qu and Perron (2008) considered methods to estimate the following model for demeaned daily returns r_t :

$$r_t = \exp(h_t/2 + \mu_t/2)\varepsilon_t, \quad h_{t+1} = \phi h_t + \sigma_v v_t, \quad \mu_{t+1} = \mu_t + \delta_t \sigma_\eta \eta_t,$$

where η_t, v_t and ε_t are independent $N(0,1)$ variables, δ_t is a sequence of independent Bernoulli random variables taking value 1 with unknown probability p (i.e., $\delta_t \sim B(1,p)$) and $\eta_j, \delta_h, \varepsilon_k$ and v_l are mutually independent for all $1 \leq j, h, k, l \leq n$. The model implies

$$\log r_t^2 = h_t + \mu_t + \log \varepsilon_t^2. \tag{3}$$

Hence $\log r_t^2$ is a level shift process satisfying DGP-1. They estimated the model using S&P 500 daily returns for the period 1980.1-2005.12 (6,572 observations) adopting a Bayesian approach. The posterior means of the parameters were $\phi = 0.953$, $\sigma_v = 0.148$, $\sigma_\eta = 1.679$, $\delta_t \sim B(1, 0.00187)$ and $(h_1, \mu_1) = (0.450, -0.210)$. These imply a level shift every 535

days, on average, whose magnitude is however large, their standard deviation being 1.679, compared to a standard deviation of 1.047 for the return series during the period considered.

Using these estimates, we generated 500 samples of size $T = 8,192$ (2^{13}) using (3). For each simulated sample, we computed the autocorrelations for all lags between 1 and 8,192 and the log-periodogram estimates of the long-memory parameter using a wide range for the number of frequency ordinates. The averages over the 500 samples are reported in Figure 9. They generate patterns that are in close agreement with the theory and the empirical estimates. Hence, the main features uncovered by our theoretical results are easily explained by a simple level shift model with a short-memory component that is of empirical relevance.

5 Can a long memory process explain the empirical features?

We now consider whether the features of the data analyzed can be explained by various models with long-memory, with or without a level shift component. To do so we resort to selected simulation experiments. We performed extensive simulations and only report ones that are representative of the problem, paying special attention to processes calibrated to actual data, in particular the S&P 500 returns series over the period 1973-2002, which as we shall see is the least favorable for the level shift model. The basic data generating process is given by the following ARFIMA model with level shifts:

$$x = c + u_{T,t} + v_t, \quad (1 - \alpha L)(1 - L)^d v_t = (1 - \theta L)e_t, \quad (4)$$

where $e_t \sim i.i.d. N(0, \sigma_e^2)$ and $u_{T,t}$ is as defined in (1). Without loss of generality, we set $c = 0$. In all cases, we use a sample of $T = 8,192$ (2^{13}) observations (roughly of the same order as a daily series over the period 1973-2002) and all results are based on 1,000 replications.

Consider first the simple case of a pure long-memory process with $u_{T,t} = \alpha = \theta = 0$ and $d = 0.457$, a typical value in the volatility literature. Figure 10 presents the means of the autocorrelations and the log-periodogram estimates of d as a function of m , along with 90% confidence intervals. The autocorrelation function has a pattern close to that of a short-memory process with level shifts though, on average, it decreases faster and crosses the zero line earlier at roughly 15% of the sample with the value of d considered. From unreported simulations, the decrease to zero is faster as d is smaller (even with $d = 0.8$, it crosses the zero axis at only 20% of the sample). This strong bias is due to the fact that the mean of the series is imprecisely estimated (see. e.g., Percival, 1993). With respect to the path of the log-periodogram estimates as a function of m , the picture is very different. The mean is flat at the true value. Hence, a pure long memory process cannot explain the features of

the data. Consider now adding a level shift component to the pure long-memory process. The values used are $p = 12$ and $\sigma_v = 1.679$ and are based on estimates reported in Qu and Perron (2008) for the S&P 500 series over the period 1980-2005. The results are presented in Figure 11. The mean values of the autocorrelations now more closely match those of a short-memory process with level shifts. However, the path of the log-periodogram estimates is clearly different. When $m = T^{1/3}$, the mean is indeed quite high at (roughly) 0.83 but the decrease when m reaches $T^{1/2}$ is much smaller, the mean value being 0.7, well above the typical value of 0.45 found in the literature. Also, the decrease as m increases further is quite small and the estimates stay above 0.5 for all values of m .

We now consider the more realistic case of a long memory process with an important short-memory component. As found in the literature, an ARFIMA(1, d ,1) provides a reasonable fit to the data. The estimates for the S&P 500 returns series over the period 1973-2002 are $\hat{d} = 0.457$, $\hat{\alpha} = 0.298$ and $\hat{\theta} = 0.751$. Hence, there is a strong mean reverting behavior given the large negative MA coefficient for the short-memory component. The results for this process without level shifts are presented in Figure 12. Consider first the autocorrelation function. Compared to a short-memory process with level shifts, it decreases to zero much faster and the mean values are close to zero for lags beyond 1,000 (roughly 12% of the sample size). Again, the rate of decrease is faster and the point at which the autocorrelations cross the zero axis is earlier with smaller values of d . More importantly, the means of the log-periodogram estimates of d show a different pattern. The mean value when $m = T^{1/3}$ is roughly the true value, well below one, though the 90% confidence interval is quite wide. Also, when m reaches $T^{1/2}$, the decrease is small, the mean value being roughly 0.45. Of importance, however, is the substantial decline in the estimates of d when m increases further. This is due to the strongly mean reverting short-memory component and will have to be taken into consideration when constructing tests in the next section.

Consider now adding a level shift component to this ARFIMA(1, d ,1) process with the same values as specified above. The results are presented in Figure 13. The mean of the autocorrelation function is now indeed close to that of a short-memory process with level shifts. The pattern of the log periodogram estimates of d is, however, clearly different. When $m = T^{1/3}$, the mean is indeed close to 1, but the decrease when m reaches $T^{1/2}$ is small, the mean being roughly 0.9 with the lower bound of the 95% confidence interval at (roughly) 0.78. Hence, such a process cannot explain the sharp decline observed in the data with a value close to one when $m = T^{1/3}$ and a value below 0.5 when $m = T^{1/2}$. When m is increased beyond $T^{1/2}$ there is a steady decrease in the estimates but the values stay well

above 0.5, unlike what is found in the data analyzed.

We performed a variety of experiments with different configurations and the qualitative results are the same. In our simulations, we did not consider DGPs with fat-tailed distributions. Doing so would only reinforce the conclusion that long-memory processes are unable to replicate the features of the path of the log-periodogram estimates. Indeed, as shown by the simulations of Wright (2002), the presence of outliers or fat-tailed errors exacerbates the negative bias in the log-periodogram estimate of d for all values of m . Hence, a model with fat-tailed errors would be even less able to generate estimates of d above 0.5 for small values of m . The intuition is fairly straightforward given that outliers acts as a strong mean-reverting component which makes a series look more like a short-memory process.

To summarize, the findings of importance are: 1) in the absence of level shifts, the log-periodogram estimates of d are not above 0.5 or near one when m is small and there is accordingly no discontinuity between the estimates with $m = T^{1/3}$ and $m = T^{1/2}$; 2) with a level shift component, the estimates of d are high when m is small but, again, there is no sharp decrease, only a gradual decrease with the estimate staying above 0.5; 3) given the large confidence interval in the estimates of d when m is small, a long-memory process with a strongly mean reverting short-memory component could potentially account for the decline when m is small as well as the decrease when m is increased further. These considerations will be used in devising a testing strategy in the next section.

6 A simple test of long memory against mean shifts

Motivated by the results of Perron and Qu (2007) and the simulations reported above, we propose simple tests of the null hypothesis of a stationary long memory process designed to have power against a short-memory process affected by mean shifts. The test is based on the fact that in the latter case the estimate of d crucially depends on the number of frequencies included in the log-periodogram regression, in particular on the fact that its limit is different when a different proportion of the sample size is used for the number of frequencies. On the other hand, if the true process is a fractionally integrated one, the limiting distribution of the log periodogram estimate is the same for a wide range of the number of frequencies used.

Let $\hat{d}_{a,c}$ denote the log periodogram estimate of the memory parameter when $m_{a,c} = c[T^a]$ frequencies are included in the regression. Under the null hypothesis of a stationary fractionally integrated process, we have given some conditions on d and m (Horvich, Deo and Brodsky, 1998, Deo and Hurvich, 2001), $\sqrt{c[T^a]}(\hat{d}_{a,c} - d_0) \rightarrow^d N(0, \pi^2/24)$. Now, let

$0 < a < b < 1$ with $a < 4/5$, the test statistic proposed is simply

$$t_d(a, c_1; b, c_2) = \sqrt{24c_1 [T^a] / \pi^2} (\hat{d}_{a,c_1} - \hat{d}_{b,c_2}).$$

This test has a limiting $N(0, 1)$ distribution under the null hypothesis since

$$\sqrt{24c_1 [T^a] / \pi^2} (\hat{d}_{a,c_1} - \hat{d}_{b,c_2}) = \sqrt{24c_1 [T^a] / \pi^2} (\hat{d}_{a,c_1} - d_0) - \sqrt{24c_1 [T^a] / \pi^2} (\hat{d}_{b,c_2} - d_0)$$

and $\sqrt{24c_1 [T^a] / \pi^2} (\hat{d}_{a,c_1} - d_0) \rightarrow^d N(0, 1)$, while $\sqrt{24c_1 [T^a] / \pi^2} (\hat{d}_{b,c_2} - d_0) = \sqrt{24 [T^b] / \pi^2} (\hat{d}_{b,c_2} - d_0) \sqrt{c_1 [T^a] / [T^b]} \rightarrow 0$. It also diverges to $+\infty$ under the alternative hypothesis of a stationary process affected by level shifts since the limits of \hat{d}_{a,c_1} and \hat{d}_{b,c_2} are different and the limit of \hat{d}_{b,c_2} is lower than that of \hat{d}_{a,c_1} . These properties are stated in the following Proposition.

Proposition 4 *Suppose the series $\{x_t\}$ is a stationary Gaussian process with spectral density in a neighborhood of zero given by $f_x(w) = |1 - \exp(-iw)|^{-2d} f^*(w)$, with the function $f^*(w)$ satisfying $f^*(0) > 0$, $f^{*'}(0) = 0$, $|f^{*''}(w)| < B_1 < \infty$ and $|f^{*'''}(w)| < B_2 < \infty$ for some finite B_1, B_2 . If $0 < a < b < 1$ and $a < 4/5$, then*

$$t_d(a, c_1; b, c_2) = \sqrt{24c_1 [T^a] / \pi^2} (\hat{d}_{a,c_1} - \hat{d}_{b,c_2}) \rightarrow^d N(0, 1)$$

Note that the set of assumptions used follow Horvich, Deo and Brodsky (1998). While the Gaussian assumption is restrictive it appears to be common in the long-memory literature. It can, however, be relaxed as shown by Deo and Hurvich (2001). They consider a stochastic volatility model where the log-squared errors, say Z_t , is generated by $Z_t = \mu + Y_t + u_t$, where u_t is a short-memory process allowed to be non-Gaussian and Y_t is a Gaussian long-memory process. The result stated in Proposition 4 remains valid in this case under more stringent conditions on the rate of increase of m in relation to d (see their Theorem 2).

We use this to devise testing procedures to distinguish between a short-memory process with level shifts and a long memory process with or without level shifts taking into considerations the key features documented in the simulations presented in the last section. The first issue is to assess whether there is a steady decline in the log-periodogram estimates for values of m greater than $T^{1/2}$, which is consistent with a short-memory process with level shifts but not with a pure long memory process without level shifts. To that effect we simply use the statistic $t_d(1/2, 1; 4/5, 1)$. The second issue is to assess whether there is a sharp decline in the estimate of d when m varies between the range $T^{1/3}$ and $T^{1/2}$. There are two problems in this case. First, the maximal value need not occur at exactly $T^{1/3}$. For this reason, we consider two statistics that will allow some flexibility. The first

is the maximum of $t_d(1/3, c_1; b, 1)$ with b fixed at $1/2$ and varying c_1 between 1 and 2, i.e., $\text{sup-}t_d = \sup_{c_1 \in [1, 2]} t_d(1/3, c_1; b, 1)$. We also consider the mean value over the same range, i.e., $\text{mean-}t_d = \text{mean}_{c_1 \in [1, 2]} t_d(1/3, c_1; b, 1)$. We tried different ranges and the results reported below did not change. Second, the limit distributions of the $\text{sup-}t_d$ and $\text{mean-}t_d$ tests are not available and the finite sample distribution is likely to be affected by the underlying DGP under the null hypothesis given the simulation results presented in the last section. To overcome these problems, we use a parametric bootstrap procedure to compute the relevant critical values. For a given series, we estimate an ARFIMA(1, d , 1) model, which was shown to provide a good fit to the data (e.g., Stărică and Granger, 2005). We then simulate the null distribution of the tests using this as the DGP.

The results for all the series considered are presented in Table 1. Note first that the statistic $t_d(1/2, 1; 4/5; 1)$ is significant at the 1% level for all series, consistent with a short-memory process with level shifts (though also consistent with a long-memory process having a strongly mean reverting component). Of more importance are the results for the tests $\text{sup-}t_d$ and $\text{mean-}t_d$. For the S&P 500 full sample series, both tests reject at the 1% level. For the sub-sample 1957-2002, they reject at the 5% level. There is no rejection though for the sample 1973-2002 but a rejection at the 5% level using the $\text{sup-}t_d$ and at the 10% level using the $\text{mean-}t_d$ for the period 1990-2002, despite the fact that it has a relatively small sample size. For the NASDAQ and AMEX indices, the $\text{mean-}t_d$ test rejects at the 5% level, while the $\text{sup-}t_d$ test rejects at the 10% level. For the Dow Jones index, we have a rejection at the 1% level with the $\text{mean-}t_d$ and at the 5% level with the $\text{sup-}t_d$.

Overall, the results are consistent with a short-memory process with level shifts: a) there is significant sharp decrease in the estimate of d for small values of m ; b) there is a further decrease when m is large with estimates well below 0.5 in this range, which is not expected with a long-memory process with level shifts, in which case the estimates do decline but stay above 0.5, unless the memory parameter is small and the short-memory noise is large.

To assess whether the rejections obtained can be explained by finite sample biases occurring when the underlying process is one of long memory possibly contaminated by noise, fat-tailed errors and/or outliers, we conducted simulations. These are calibrated to empirical estimates obtained from the log squared S&P 500 returns series over the period 1973-2002. We first estimated an ARFIMA(1, d , 1) process of the form (4) without level shifts. The point estimates were $\hat{d} = 0.457$, $\hat{\alpha} = 0.298$ and $\hat{\theta} = 0.751$ which are used as the parameters of the DGP for the simulations. In order to account for the potential presence of noise, fat-tailed errors and outliers, the series for each simulation are constructed using the estimated residu-

als drawn randomly with replacement. For each simulated draw, the bootstrap-based critical values were computed using 1000 simulations. The rejection frequencies at the 5% nominal level were computed from 500 such replications. We considered samples of size $T = 2^{10}$, 2^{11} , 2^{12} and 2^{13} . The results presented in Table 2 show the exact sizes of our tests to be close to 5%. In fact, the tests tend to be conservative for the larger sample sizes. This can be explained by the presence of outliers in the estimated residuals from which we draw, which act as strongly mean-reverting components and, accordingly, produce estimates of d that are further downward-biased so that the test rejects less often. Hence, we are confident that the rejections documented above are indicative of a level shift process and not the outcome of a long-memory process possibly contaminated by noise, fat-tailed errors and outliers.

7 Conclusions

Our paper provided an analysis of various statistics when the underlying model is a short-memory process contaminated by level shifts. Our theoretical results provide clear practical implications that can be confronted with the data by looking at the periodogram, the autocorrelation function and the path of the log periodogram estimates as the number of frequencies used varies. Using data on stock market volatility proxies for various indices and sample periods, our results show that a short-memory process with level shifts should be viewed as a serious contender to model volatility. All estimates considered clearly follows a pattern that would obtain if the true underlying process was one of short-memory contaminated by level shifts. Our results suggest that research should be oriented in the direction of such a class of models to understand the time series properties of stock returns volatility. Qu and Perron (2008) consider a stochastic volatility model with both a level shift and a short-memory component and present a Bayesian inference procedure. They show that the model provides a good fit to the data and forecasts as well, and better in some cases, as other leading volatility models when applied to S&P 500 and NASDAQ daily returns. Lu and Perron (2008) present a method to directly estimate a level shift model using an extension of the Kalman filter and apply it to the logarithm of squared returns for the S&P 500, AMEX, Dow Jones and NASDAQ stock market return indices. The point estimates imply few level shifts for all series. But once these are taken into account, there is little evidence of serial correlation in the remaining noise and, hence, no evidence of long-memory. They also produce rolling out-of-sample forecasts of squared returns. In most cases, the simple random level shifts model clearly outperforms a standard GARCH(1,1) model and, in many cases, it also provides better forecasts than a fractionally integrated GARCH model.

Appendix

We first state some Lemma that will be used in subsequent proofs.

Lemma A.1 (Georgiev, 2002): Let $\int X(s)dY(s)$ denote $\int X(s-)dY(s)$, with $X(s-)$ the left limit of X at s , and the stochastic integral being of the Ito type. Also, let $f(s)$ be a continuous function on $[0, 1]$. Then: 1) $T^{-1} \sum_{t=1}^{[Tr]} u_{T,t} f(t/T) \Rightarrow \int_0^r J(s) f(s) ds$; 2) $T^{-1} \sum_{t=1}^{[Tr]} u_{T,t}^2 f(t/T) \Rightarrow \int_0^r J^2(s) f(s) ds$; 3) $T^{-1/2} \sum_{t=1}^{[Tr]} u_{T,t} \varepsilon_t \Rightarrow \int_0^r J(s) dW(s)$ if ε_t is i.i.d. $(0, 1)$ independent of $u_{T,t}$.

The following lemma concerns the moments of the compound Poisson process $J(s)$. The proofs are straightforward and, hence, omitted.

Lemma A.2 a) $E[J(s)^2] = p\sigma_\eta^2 s$; and under Normality: b) $E[J(s)^4] = 3ps\sigma_\eta^4 + 3p^2 s^2 \sigma_\eta^4$; c) $E[J^2(s)(J(u) - J(s))^2] = s(u - s)p^2\sigma_\eta^4$ for $u > s$; and $E[J^2(u)(J(s) - J(u))^2] = u(s - u)p^2\sigma_\eta^4$ for $s > u$.

Proof of Propositions 1 and 2: We have

$$\hat{R}(h) = T^{-1} \sum_{t=1}^{T-h} (x_t - \bar{x})(x_{t+h} - \bar{x}) = T^{-1} \sum_{t=1}^{T-h} x_t x_{t+h} - (T^{-1} \sum_{t=1}^T x_t)^2 + o_p(1).$$

Consider the first term,

$$\begin{aligned} T^{-1} \sum_{t=1}^{T-h} x_t x_{t+h} &= T^{-1} \sum_{t=1}^{T-h} (v_t + u_{T,t}) (v_t + u_{T,t+h}) \\ &= T^{-1} \sum_{t=1}^{T-h} v_t v_{t+h} + T^{-1} \sum_{t=1}^{T-h} v_t u_{T,t+h} + T^{-1} \sum_{t=1}^{T-h} u_{T,t} u_{T,t+h} + T^{-1} \sum_{t=1}^{T-h} u_{T,t} v_{t+h} \end{aligned}$$

Now, for $h/T \rightarrow 0$ as $T \rightarrow \infty$, we have $T^{-1} \sum_{t=1}^{T-h} v_t v_{t+h} \rightarrow^p R_v(h)$, $T^{-1} \sum_{t=1}^{T-h} v_t u_{T,t+h} \rightarrow^p 0$, $T^{-1} \sum_{t=1}^{T-h} u_{T,t} v_{t+h} \rightarrow^p 0$, and $T^{-1} \sum_{t=1}^{T-h} u_{T,t} u_{T,t+h} \Rightarrow \int_0^1 J^2(s) ds$ using Lemma A.1. Similarly, $T^{-1} \sum_{t=1}^T x_t = T^{-1} \sum_{t=1}^T v_t + T^{-1} \sum_{t=1}^T u_{T,t} \Rightarrow \int_0^1 J(s) ds$. Collecting terms, the result of Proposition 1 (a) follows. When $h/T \rightarrow \kappa$ as $T \rightarrow \infty$, we have $T^{-1} \sum_{t=1}^{T-h} v_t v_{t+h} \rightarrow^p 0$ since $h \rightarrow \infty$ and the process is short memory. Also, $T^{-1} \sum_{t=1}^{T-[T\kappa]} u_{T,t} u_{T,t+[T\kappa]} \Rightarrow \int_0^1 J(s) J(s+\kappa) ds$ and the limit of the other terms remains the same. Consider now $E(R(\kappa))$. We have

$$\begin{aligned} E(R(\kappa)) &= \int_0^{1-\kappa} E((J(u) - \bar{J})(J(u+\kappa) - \bar{J})) du \\ &= \int_0^{1-\kappa} E(J(u) J(u+\kappa)) du - \int_0^{1-\kappa} E(J(u) \bar{J}) du \\ &\quad - \int_0^{1-\kappa} E(J(u+\kappa) \bar{J}) du + \int_0^{1-\kappa} E(\bar{J}^2) du \end{aligned}$$

The solutions to each each of the four terms are

$$\begin{aligned} \int_0^{1-\kappa} E(J(u) J(u+\kappa)) du &= \int_0^{1-\kappa} E(J(u)^2) du + \int_0^{1-\kappa} E(J(u) (J(u+\kappa) - J(u))) du \\ &= \int_0^{1-\kappa} E(J(u)^2) du = \int_0^{1-\kappa} pu\sigma^2 du = (1-\kappa)^2(p\sigma^2/2), \end{aligned}$$

$$\begin{aligned}\int_0^{1-\kappa} E(J(u) \bar{J}) du &= p\sigma^2 \int_0^{1-\kappa} u(1 - \frac{u}{2}) du = [\frac{(1-\kappa)^2}{2} - \frac{(1-\kappa)^3}{6}] p\sigma^2, \\ \int_0^{1-\kappa} E(J(u+\kappa) \bar{J}) du &= p\sigma^2 \int_0^{1-\kappa} (u+\kappa)(1 - (u+\kappa)/2) du = [(1-\kappa^2)/2 - (1-\kappa^3)/6] p\sigma^2 \\ \int_0^{1-\kappa} E(\bar{J}^2) du &= \int_0^{1-\kappa} (1/3) p\sigma^2 du = [(1-\kappa)/3] p\sigma^2\end{aligned}$$

and the result follows adding all the solutions.

Proof of Proposition 3: Part (1) is a standard result for stationary short memory processes. For Part (2), we have

$$\begin{aligned}\frac{j^2}{2\pi T^2} \left| \sum_{t=1}^T u_{T,t} \exp(i\frac{2\pi jt}{T}) \right|^2 &= \frac{j^2}{2\pi T^2} \sum_{s=1}^T \sum_{t=1}^T u_{T,s} u_{T,t} \cos(\frac{2\pi j(t-s)}{T}) \\ &\Rightarrow \frac{j^2}{2\pi} \int_0^1 \int_0^1 J(u) J(s) \cos(2\pi j(s-u)) ds du\end{aligned}$$

and

$$\begin{aligned}&E[\frac{j^2}{2\pi} \int_0^1 \int_0^1 J(u) J(s) \cos(2\pi j(s-u)) ds du] \\ &= E[\frac{j^2}{2\pi} \int_0^1 \int_0^u J(u) J(s) \cos(2\pi j(s-u)) ds du] + E[\frac{j^2}{2\pi} \int_0^1 \int_u^1 J(u) J(s) \cos(2\pi j(s-u)) ds du] \\ &= \frac{j^2}{2\pi} \int_0^1 \int_0^u E[J^2(s)] \cos(2\pi j(s-u)) ds du + \frac{j^2}{2\pi} \int_0^1 \int_u^1 E[J^2(u)] \cos(2\pi j(s-u)) ds du \\ &= \frac{j^2}{2\pi} \int_0^1 \int_0^u p s \sigma_\eta^2 \cos(2\pi j(s-u)) ds du + \frac{j^2}{2\pi} \int_0^1 \int_u^1 p u \sigma_\eta^2 \cos(2\pi j(s-u)) ds du = \frac{p\sigma_\eta^2}{4\pi^3},\end{aligned}$$

where we have used the fact that $E[J(u)J(s)] = E[J(\min(s,u))^2]$. The result that the variance is independent of j follows similarly from tedious algebra. For part (3), we have

$$\frac{j}{\pi T^{3/2}} \sum_{t=1}^T \sum_{s=1}^T v_t u_{T,s} \cos w_j(t-s) \Rightarrow \frac{j}{\pi} \int_0^1 \int_0^1 J(u) \cos(2\pi j(u-v)) du dW(v)$$

and the limit term is easily seen to have mean zero (it also has mean zero in finite samples since v_t and $u_{T,t}$ are assumed to be independent). For the variance, we have:

$$\begin{aligned}E[\frac{j}{\pi} \int_0^1 \int_0^1 J(u) \cos(2\pi j(u-v)) du dW(v)]^2 &= \frac{j^2}{\pi^2} \int_0^1 E[\int_0^1 J(u) \cos(2\pi j(u-v)) du]^2 dv \\ &= \frac{j^2}{\pi^2} \int_0^1 E[\int_0^1 \int_0^1 J(u) J(s) \cos(2\pi j(u-v)) \cos(2\pi j(s-v)) ds du] dv \\ &= \frac{j^2}{\pi^2} \int_0^1 E[\int_0^1 \int_0^u J(u) J(s) \cos(2\pi j(u-v)) \cos(2\pi j(s-v)) ds du] dv \\ &\quad + \frac{j^2}{\pi^2} \int_0^1 E[\int_0^1 \int_u^1 J(u) J(s) \cos(2\pi j(u-v)) \cos(2\pi j(s-v)) ds du] dv \\ &= \frac{j^2}{\pi^2} \int_0^1 \int_0^1 \int_0^u p^2 s \sigma_\eta^4 \cos(2\pi j(u-v)) \cos(2\pi j(s-v)) ds du dv \\ &\quad + \frac{j^2}{\pi^2} \int_0^1 \int_0^1 \int_u^1 p^2 u \sigma_\eta^4 \cos(2\pi j(u-v)) \cos(2\pi j(s-v)) ds du dv = \frac{p^2 \sigma_\eta^4}{4\pi^4}.\end{aligned}$$

References

- Andersen, T.G., and Bollerslev, T. (1997), "Heterogeneous Information Arrivals and Return Volatility Dynamics: Uncovering the Long-run in High Frequency Data," *Journal of Finance*, 52, 975-1005.
- Andersen, T.G., Bollerslev, T., Deibold, F.X., and Labys, P. (2003), "Modeling and Forecasting Realized Volatility," *Econometrica*, 71, 579-625.
- Andrews, D.W.K. and Guggenberger, P. (2003), "A Bias-reduced Log-periodogram Regression Estimator for the Long-memory Parameter," *Econometrica*, 71, 675-712.
- Beran, J. (1994), *Statistics for Long Memory Processes*, New York: Chapman & Hall.
- Baillie, R.T. (1996), "Long Memory Processes and Fractional Integration in Econometrics," *Journal of Econometrics*, 73, 5-59.
- Bollerslev, T., and Wright, J.H. (2000), "Semiparametric Estimation of Long-memory Volatility Dependencies: The Role of High Frequency Data," *Journal of Econometrics*, 98, 81-106.
- Breidt, F.J., Crato, N., and de Lima, P. (1998), "The Detection and Estimation of Long Memory in Stochastic Volatility," *Journal of Econometrics*, 83, 325-348.
- Deo, R.S., and Hurvich, C.M. (2001), "On the Log Periodogram Regression Estimator of the Memory Parameter in Long Memory Stochastic Volatility Models," *Econometric Theory*, 17, 686-710.
- Diebold, F., and Inoue, A. (2001), "Long Memory and Regime Switching," *Journal of Econometrics*, 105, 131-159.
- Ding, Z., Engle, R.F., and Granger, C.W.J. (1993), "A Long Memory Property of Stock Market Returns and a New Model," *Journal of Empirical Finance*, 1, 83-106.
- Engle, R.F., and Smith, A.D. (1999), "Stochastic Permanent Breaks," *Review of Economics and Statistics*, 81, 533-574.
- Georgiev, I. (2002), "Functional Weak Limit Theory for Rare Outlying Events," Unpublished Manuscript, European University Institute.
- Geweke, J., and Porter-Hudak, S. (1983) "The Estimation and Application of Long Memory Time Series Models," *Journal of Time Series Analysis*, 4, 221-238.
- Gourieroux, C., and Jasiak, J. (2001), "Memory and Infrequent Breaks," *Economics Letters*, 70, 29-41.
- Granger, C.W.J. (1981), "Some Properties of Time Series Data and Their Use in Econometric Model Specification," *Journal of Econometrics*, 16, 121-130.

- Granger, C.W.J., and Ding, Z. (1995), "Some Properties of Absolute Return: an Alternative Measure of Risk," *Annales d'Économie et de Statistique*, 40, 67-91.
- Granger, C.W.J., and Ding, Z. (1996), "Varieties of Long Memory Models," *Journal of Econometrics*, 73, 61-77.
- Granger, C.W.J., and Hyung N. (2004), "Occasional Structural Breaks and Long Memory with an Application to the S&P 500 Absolute Stock Returns," *Journal of Empirical Finance*, 11, 399-421.
- Granger, C.W.J., and Joyeux, R. (1980), "An Introduction to Long Memory Time Series Models and Fractional Differencing," *Journal of Time Series Analysis*, 1, 15-39.
- Granger, C.W.J., and Terasvirta, T. (1991), "A Simple Nonlinear Time Series Model with Misleading Linear Properties," *Economics Letters*, 62, 161-165.
- Hosking, J.R.M. (1981), "Fractional Differencing," *Biometrika*, 68, 165-176.
- Hurvich, C.M., Deo, R., and Brodsky, J. (1998), "The Mean Squared Error of Geweke and Porter-Hudak's Estimator of the Memory Parameter of a Long-memory Time Series," *Journal of Time Series Analysis*, 19, 19-46.
- Hurvich, C.M., Moulines, E., and Soulier P. (2005), "Estimating Long Memory in Volatility," *Econometrica*, 73, 1283-1328.
- Leipus, R., and Viano, M.-C. (2003), "Long Memory and Stochastic Trend," *Statistics and Probability Letters*, 61, 177-190.
- Lobato, I.N., and Savin, N.E. (1998), "Real and Spurious Long-memory Properties of Stock-market Data," *Journal of Business and Economics Statistics*, 16, 261-268.
- Lu, Y.K., and Perron, P. (2008), "Modeling and Forecasting Stock Return Volatility Using a Random Level Shift Model," Unpublished Manuscript, Department of Economics, Boston University.
- Mikosch, T., and Stărică, C. (2004a), "Nonstationarities in Financial Time Series, the Long-range Effect Dependence, and the IGARCH Effects," *Review of Economics and Statistics*, 86, 378-390.
- Mikosch, T., and Stărică, C. (2004b), "Changes of Structure in Financial Time Series and the GARCH model," *REVSTAT - Statistical Journal*, 2, 42-73.
- Parke, W.R. (1999), "What is Fractional Integration?" *Review of Economics and Statistics*, 81, 632-638.
- Percival, D.B. (1993), "Three Curious Properties of the Sample Variance and Autocovariance for Stationary Processes with Unknown Mean," *The American Statistician*, 47(4), 274-276.

- Perron, P. (1989), "The Great Crash, the Oil Price Shock and the Unit root Hypothesis," *Econometrica*, 57, 1361-1401.
- Perron, P. (1990), "Testing for a Unit Root in a Time Series Regression with a Changing Mean," *Journal of Business and Economic Statistics*, 8, 153-162.
- Perron, P. and Qu Z. (2007), "An Analytical Evaluation of the Log-periodogram Estimate in the Presence of Level Shifts," Unpublished Manuscript, Department of Economics, Boston University.
- Phillips, P.C.B. (2007), "Unit Root Log Periodogram Regression," *Journal of Econometrics*, 138, 104-124.
- Qu, Z., Perron, P. (2008), "A Stochastic Volatility Model with Random Level Shifts: Theory and Applications to S&P 500 and NASDAQ Return Indices," Unpublished Manuscript, Department of Economics, Boston University.
- Robinson, P.M. (1994), "Time Series with Strong Dependence," in: C. Sims, ed., *Advances in Econometrics*, 6th World Congress (Cambridge University Press, Cambridge), 47-95.
- Robinson, P.M. (1995), "Log Periodogram Regression of Time Series with Long Range Dependence," *Annals of Statistics*, 23, 1048-1072.
- Schwert, G.W. (1990), "Indexes of United States Stock Prices, 1802-1987," *Journal of Business*, 63, 399-426.
- Stărică, C., and Granger, C. (2005), "Nonstationarities in Stock Returns," *Review of Economics and Statistics*, 87, 503-522.
- Sun, Y., and Phillips, P.C.B. (2003), "Nonlinear Log-periodogram Regression for Perturbed Fractional Processes," *Journal of Econometrics*, 115, 355-389.
- Teverovsky, V., and Taqqu, M. (1997), "Testing for Long-range Dependence in the Presence of Shifting Means or a Slowly Declining Trend, Using a Variance-type Estimator," *Journal of Time Series Analysis*, 18, 279-304.
- Wichern, D. (1973), "The Behaviour of the Sample Autocorrelation Function for an Integrated Moving Average Process," *Biometrika*, 60, 235-239.
- Wright, J.H. (2002), "Log-periodogram Estimation of Long Memory Volatility Dependencies with Conditionally Heavy Tailed Returns," *Econometric Reviews*, 21, 397-417.

Figure 1: S&P 500 daily returns: 1928.1–2002.10

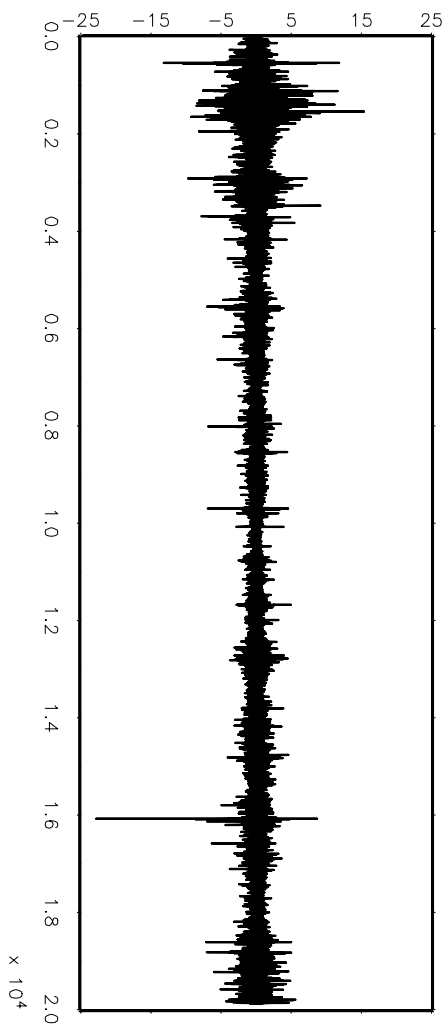
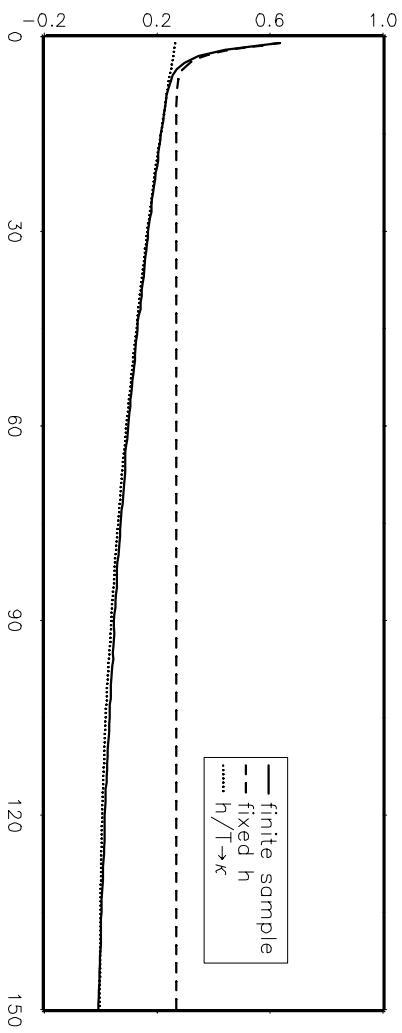


Figure 2: Median of the autocorrelations: finite sample and asymptotic approximations:

(a) $\rho=0.5$ $T=500$



(b) $\rho=0.9$ $T=500$

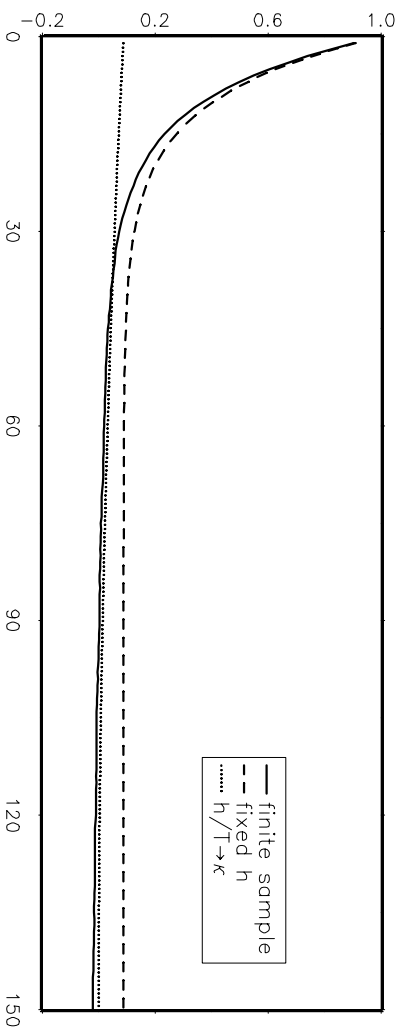
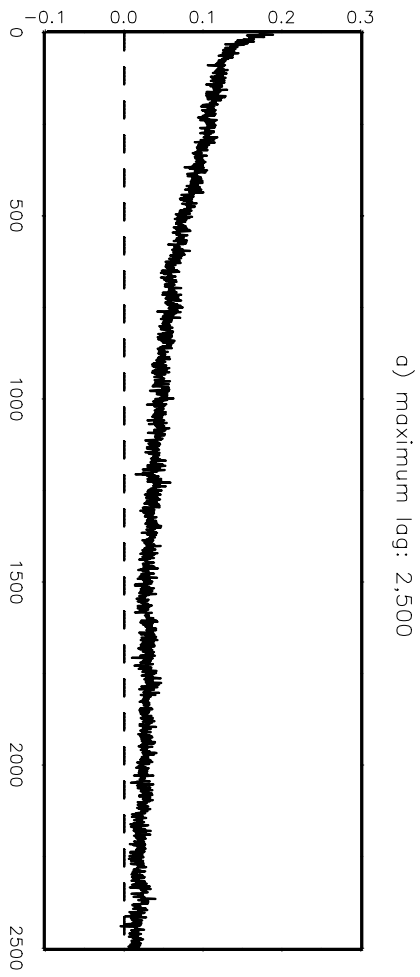


Figure 3: The autocorrelation function of S&P500 log squared returns



b) maximum lag: 19,867

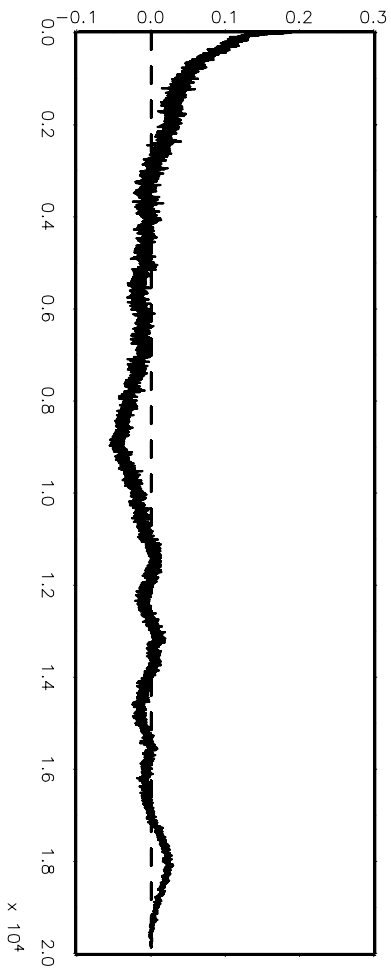
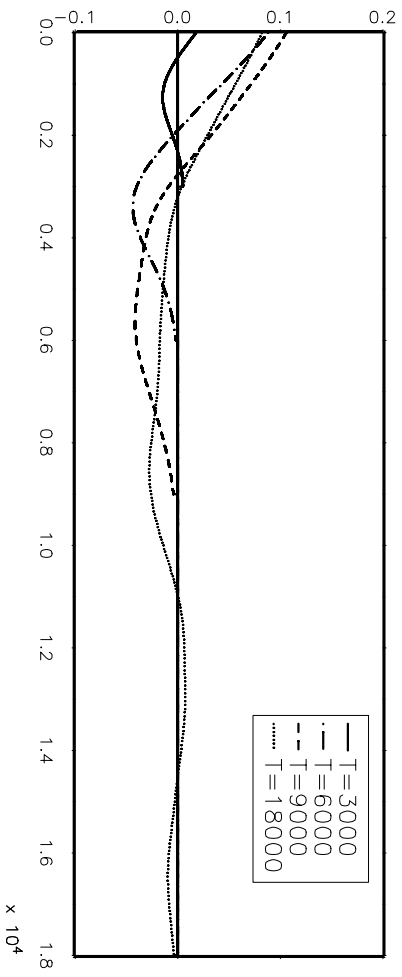


Figure 4 : The autocorrelation functions of S&P500 log squared returns with different sample sizes
[all samples start on January 4, 1928]



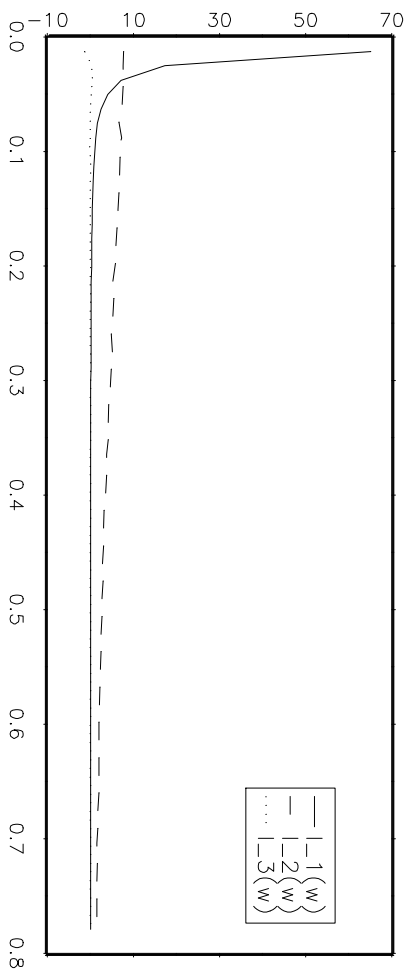


Figure 5: Median of the periodogram at frequencies 0 to $\pi/8$; $T=500$, $\rho=0.7$

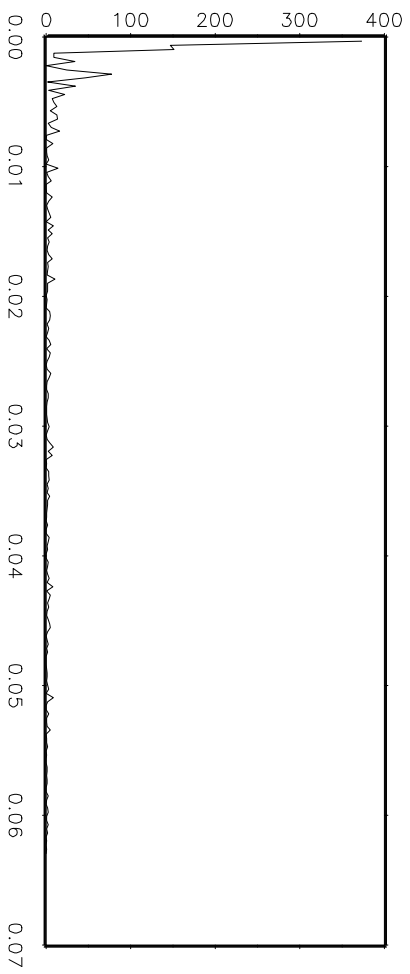


Figure 6: The periodogram at frequencies 0 to $\pi/50$ for S&P500 log squared returns

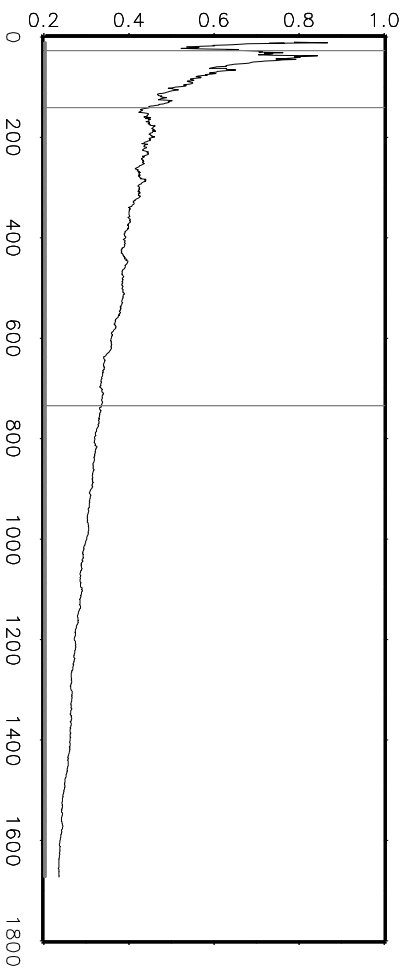


Figure 7: The estimates of d with different m for S&P500 log squared returns

Figure 8: The results for S&P500 log squared returns for the period 1990.12 to 2002.10

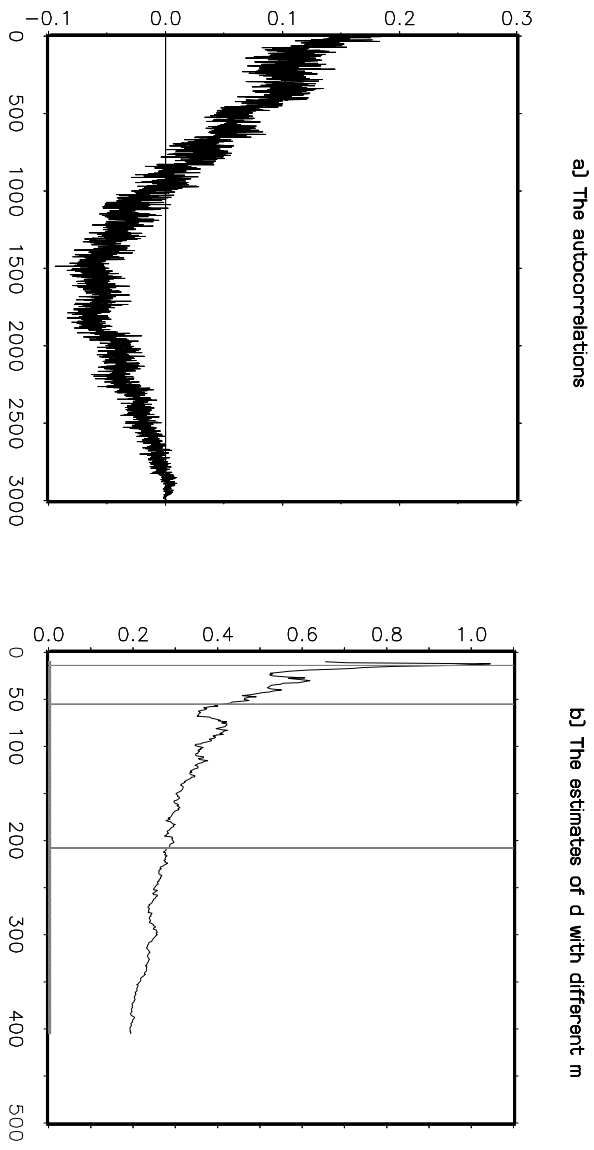


Figure 9: Spectral and time domain properties of the simulated series: $T=2^{13}$

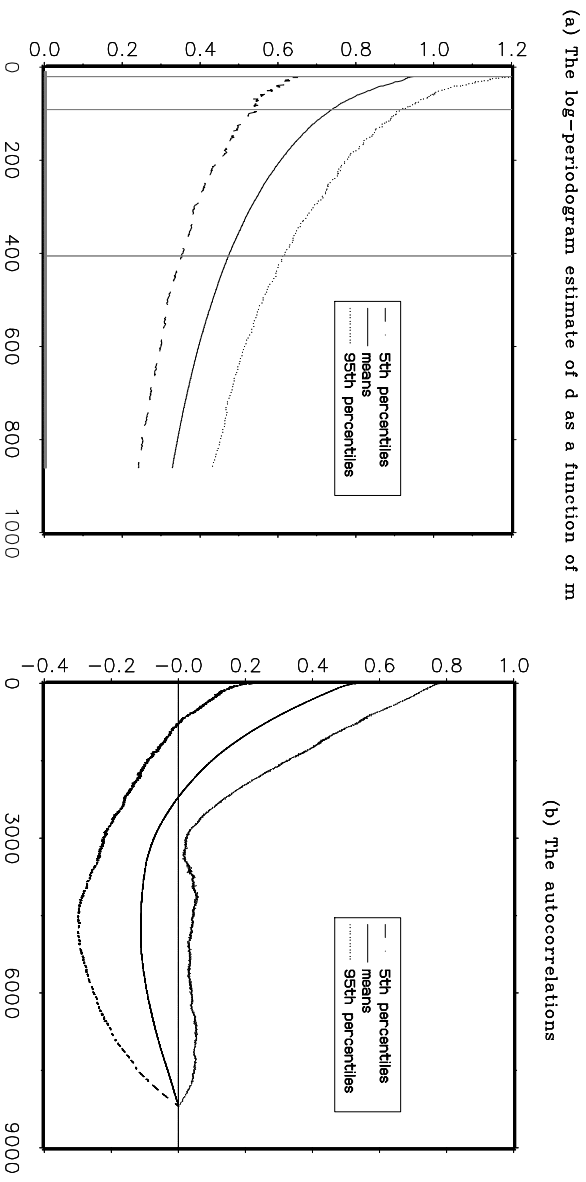


Figure 10: The results for simulated ARFIMA(0,d,0) series: $T=2^{13}$

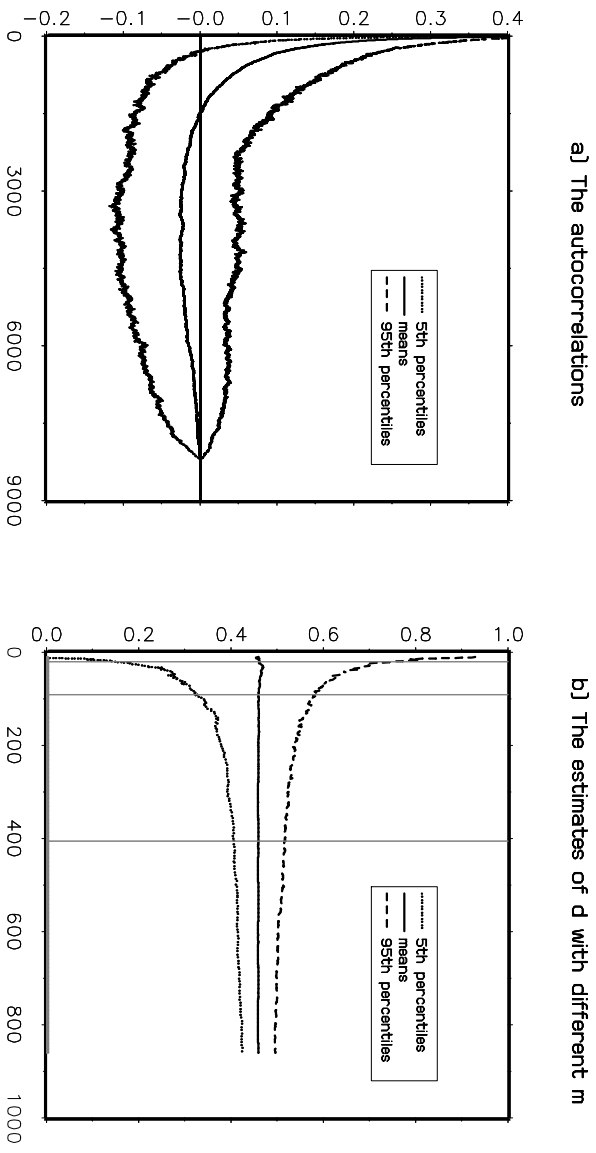


Figure 11: The results for simulated ARFIMA(0,d,0) plus level shifts series: $T=2^{13}$

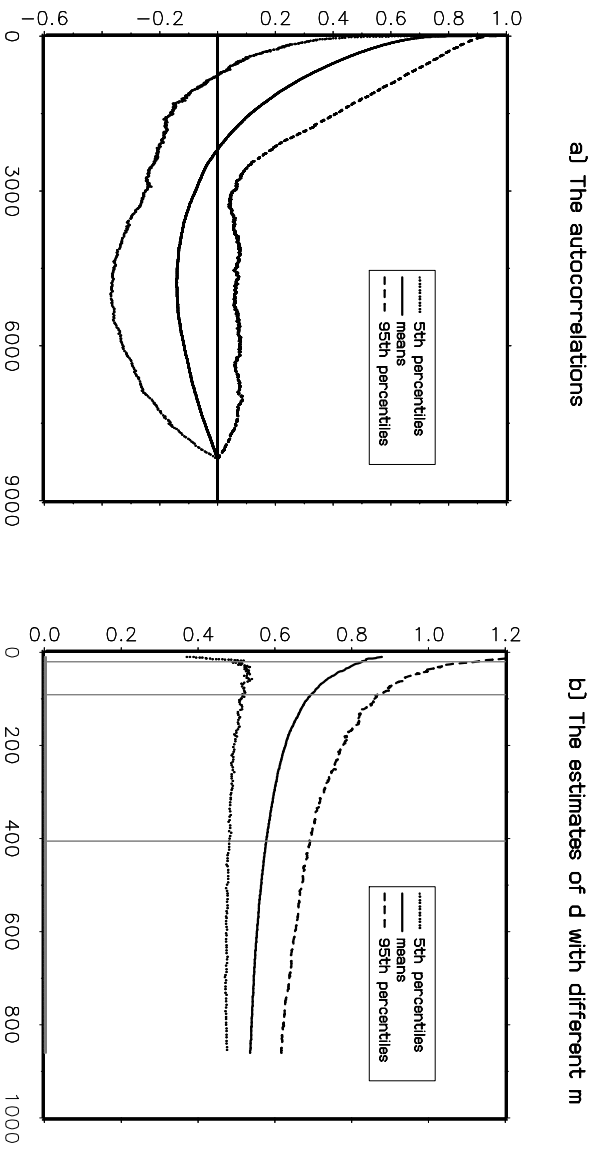


Figure 12: The results for simulated ARFIMA(1,d,1) series: $T=2^*13$

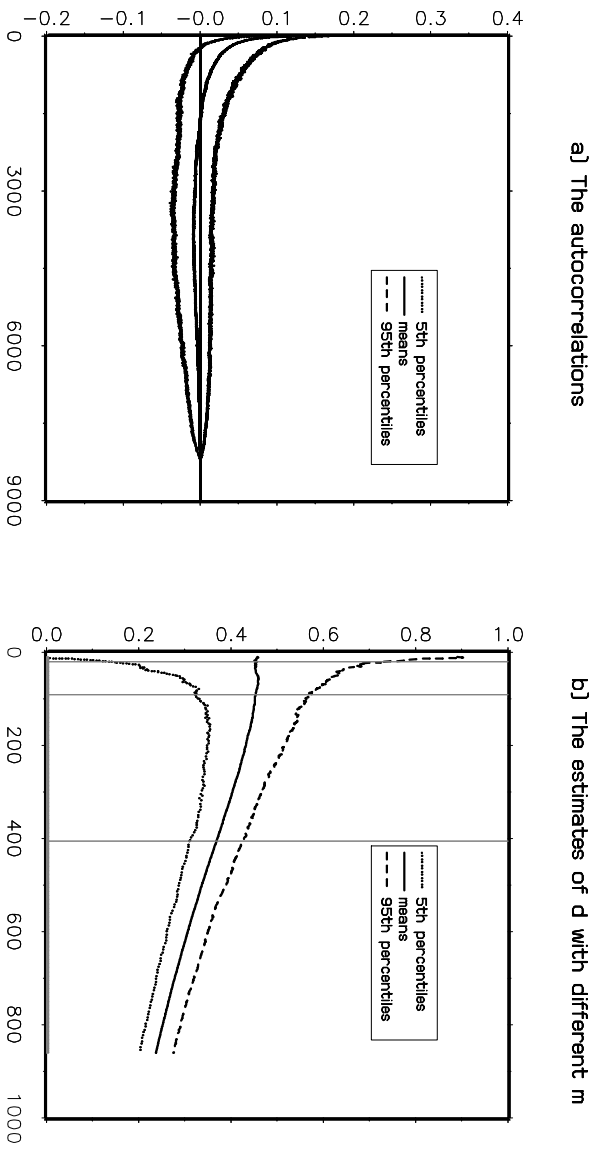


Figure 13: The results for simulated ARFIMA(1,d,1) plus level shifts series: $T=2^*13$

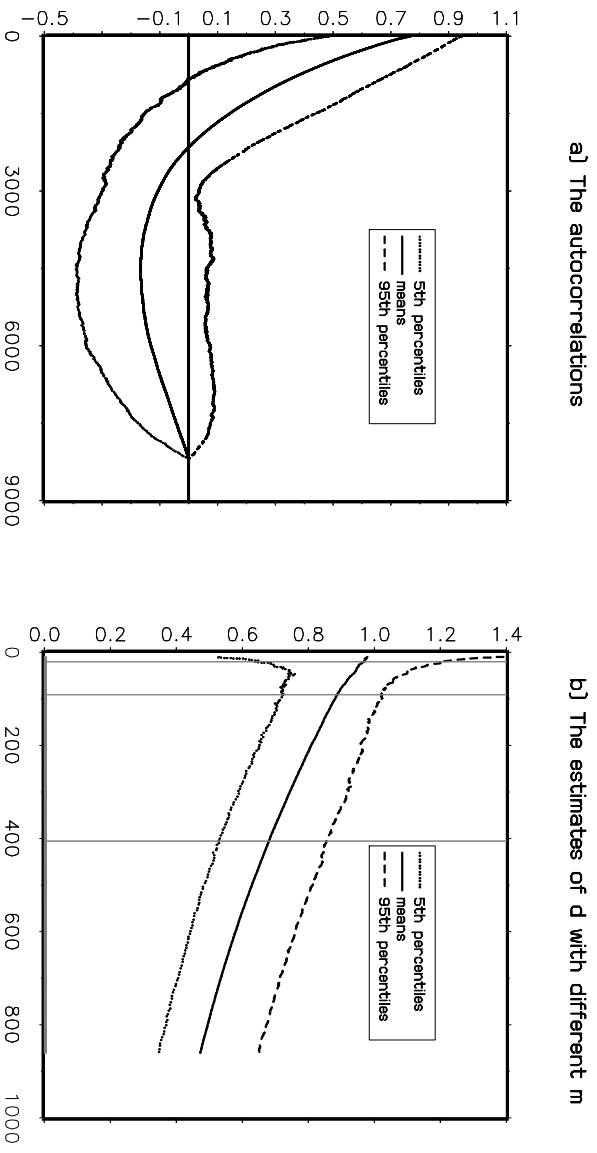


Table 1: Results of the tests

Series	Sample Period	Sample size	Tests		Simulated DGP		Simulated Critical Values	
			$t_d(\frac{1}{2}, 1, \frac{4}{5}, 1)$	Sup- t_d	Mean- t_d	ARFIMA(1,d,1)	Sup- t_d	Mean- t_d
S&P 500	1928.1-2002.10	19868	8.49	3.85	2.78	(0.31,0.47,0.73)	2.05,2.39,3.08	1.28,1.62,2.23
S&P 500	1957.3-2002.10	11494	7.55	2.72	2.05	(0.33,0.46,0.75)	2.06,2.45,2.96	1.24,1.54,2.07
S&P 500	1973.1-2002.10	7532	7.21	1.49	1.09	(0.30,0.46,0.75)	2.09,2.44,3.49	1.25,1.60,2.38
S&P 500	1990.12-2002.10	3000	4.79	3.19	1.48	(0.26,0.45,0.69)	2.23,2.63,3.69	1.26,1.67,2.46
Dow Jones	1957.3-2002.10	11534	7.87	3.17	2.39	(0.33,0.43,0.73)	2.08,2.48,3.22	1.28,1.56,2.30
AMEX	1961.7-2006.12	11201	7.34	2.36	1.71	(0.29,0.48,0.72)	2.12,2.50,3.22	1.27,1.55,2.36
NASDAQ	1972.12-2006.12	8592	7.99	2.46	1.71	(0.29,0.48,0.72)	2.17,2.62,3.47	1.28,1.67,2.44

Note: The simulated critical values correspond to 10%, 5% and 1% significance levels. $\hat{d}_{a,c}$ denotes the log periodogram estimate of d when $c[T^a]$ frequencies are included; $t_d(a, c_1; b, c_2) \equiv \sqrt{24c_1 [T^a] / \pi^2 (\hat{d}_{a,c_1} - \hat{d}_{b,c_2})}$; Sup- $t_d \equiv \sup_{c_1 \in [1,2]} t_d(1/3, c_1; 1/2, 1)$; Mean- $t_d \equiv Mean_{c_1 \in [1,2]} t_d(1/3, c_1; 1/2, 1)$.

Table 2: Size of the tests using simulated ARFIMA series

Sample size	Empirical rejection frequencies	
	Sup- t_d	Mean- t_d
$T = 2^{10}$	0.066	0.058
$T = 2^{11}$	0.050	0.046
$T = 2^{12}$	0.042	0.044
$T = 2^{13}$	0.036	0.044