TRUE vs SPURIOUS LONG MEMORY: SOME THEORETICAL RESULTS AND A MONTE CARLO COMPARISON*

Arturo Leccadito[1], Omar Rachedi[2], and Giovanni Urga[3]

[1] Università della Calabria, Department of Business Administration, Arcavacata di Rende, Italy
[2] Universidad Carlos III, Getafe, Spain
[3] Centre for Econometric Analysis, Faculty of Finance, Cass Business School, London, U.K. and Università di Bergamo, Italy

A common feature of financial time series is their strong persistence. Yet, long memory may just be the spurious effect of either structural breaks or slow switching regimes. We explore the effects of spurious long memory on the elasticity of the stock market price with respect to volatility and show how cross-sectional aggregation may generate spurious persistence in the data. We undertake an extensive Monte Carlo study to compare the performances of five tests, constructed under the null of true long memory versus the alternative of spurious long memory due to level shifts or breaks.

Keywords: Fractional integration; Structural Break; Regime Switching.

JEL Classification: C15; C22; G10.

*Address correspondence to Giovanni Urga, Centre for Econometric Analysis, Faculty of Finance, Cass Business School, 106 Bunhill Row, EC1Y 8TZ London, (U.K.); E-mail: g.urga@city.ac.uk

1. INTRODUCTION

Most financial and economic time series feature strong persistence over time. However, long memory (LM) behavior may arise as a spurious effect of the presence of either structural breaks or slow switching regimes. Several papers have attempted to disentangle LM and structural changes. Diebold and Inoue (2001) provide both theoretical justification and Monte Carlo evidence showing that a time series with structural breaks can induce a strong persistence in the autocorrelation function and hence generate spurious LM. Granger and Hyung (2004) find that an occasional-break model and an $I(d)$ model equally well explain the absolute stock returns series. Smith (2005) shows that the Geweke and Porter-Hudak (1983) (GPH) estimator, when applied to stationary short-memory processes with a slowly varying level, erroneously detects the presence of LM. Choi and Zivot (2007) find that the persistence in the forward discount series for five G7 countries is considerably reduced when adjusted for multiple breaks. Yet, the evidence of LM cannot be rejected.

Though important both in finance and economics, the issue of discriminating between genuine and spurious LM has been partly neglected by the empirical literature due to the lack of sound theoretical guidelines. The main contribution of this paper is twofold. First, we develop two theoretical results, relevant to empirical work, related to long standing topics in finance and economics. In particular, we evaluate the impact of spurious LM on a) the elasticity of stock price to volatility and b) on aggregating micro series.

In the empirical literature, the realized measure of volatility is modelled as a highly persistent stationary process (see Andersen et al., 2001, 2003). Yet, Hyung et al. (2007) claim that LM in realized volatility can arise due to spurious phenomena. Baillie and Kapetanios (2007) and McAleer and Medeiros (2008) show that nonlinearities in the LM component may spuriously increase the degree of persistence. Christensen and Nielsen

(2007) demonstrate that the presence of a LM in realized volatility, in conjunction with a positive risk-return tradeoff, entails a rather modest elasticity of stock prices with respect to volatility changes. This is a surprising result, because one would expect shocks to volatility to have a long-lasting effect on stock prices. However, Christensen and Nielsen (2007) do not consider the possibility that volatility may follow a spurious LM process. In Section 2.1, we develop a counterexample to evaluate the impact of spurious LM on volatility in the framework of Christensen and Nielsen (2007).

With respect to the aggregation issue, our analysis is inspired by the results of Granger (1980), Zaffaroni (2004) and Souza (2008), and is relevant for empirical works on inflation. In the literature there is no consensus on whether aggregate inflation, measured as a cross-sectional mean of sectoral CPI indexes, is a fractional integrated process, as shown in Altissimo et al. (2009), rather than a regime switching process, as in Benati (2008). In addition, Altissimo et al. (2009) observe low persistence at the level of sectoral inflation but high persistence in the aggregate inflation. In Section 2.2, we highlight how cross-sectional aggregation may generate spurious persistence in the data.

Given that an important step for robust inference and sound empirical applications is the identification of the genuine property of the series, the second contribution of this paper is to offer some guidelines to practitioners on the appropriate use of suitable tests. We undertake an extensive Monte Carlo study and check the overall performance of the available testing procedures to distinguish between true and spurious LM. To the best of our knowledge, five testing procedures have been proposed with the null of true stationary LM. Ohanissian et al. (2007) develop a test to distinguish between true and spurious LM based on the invariance principle of the LM parameter for temporal aggregates. Authors find that the exchange rate volatility is driven by a true LM process. Dolado et al. (2005) propose a time-domain test to verify whether a process is $I(d)$, against the alternative of being $I(0)$ with deterministic components subject to structural breaks at known or unknown dates. The test is applied to the study of U.S. real GNP and absolute values and squared values of S&P 500 returns, and in both cases the null of $I(d)$ cannot be rejected at the conventional significance levels. Shimotsu (2006) proposes two simple tests for true versus spurious LM. The first is a Wald test comparing the degree of persistence over different subsamples, whereas the second test is based on the d-th differenced series. The tests are applied to the daily realized volatility of the S&P 500 index. Despite the presence of infrequent structural breaks in the data, the results do not provide strong evidence against the hypothesis of genuine LM. Finally, Qu (2011) introduces a new test extending the results of Perron and Qu (2010) on the analysis of the components of the periodogram estimates. The test is used to study monthly temperature, monthly US inflation, exchange rates realized volatility and log squared S&P daily returns series. The null of true LM is rejected in all the cases but in the exchange rates realized volatility series.

The remainder of the paper is organized as follows. In Section 2, we show the importance of distinguishing between genuine and spurious LM in volatility; in addition, we explore how spurious LM could arise as a consequence of cross-sectional aggregation of short memory time series. Section 3 reports results of several Monte Carlo experiments involving the five statistical tests. Sizes and the powers of the tests are compared under several alternatives allowing for short memory, structural breaks and spurious LM processes. Section 4 concludes and offers some guidelines for empirical work.

2. TWO EXAMPLES OF POTENTIAL SPURIOUS LONG MEMORY

In this section, we first evaluate the impact of a spurious LM process for volatility on the results of Christensen and Nielsen (2007). Next, we consider some aggregation issues as in Altissimo et al. (2009) and Davidson and Sibbertsen (2005), and we show that it is possible to end up with a spurious LM aggregate series, even if the single cross-sectional units do not exhibit (spurious) LM.

2.1. The Effect of Spurious Long Memory in Volatility on Stock Prices

Poterba and Summers (1986) study the elasticity of asset values with respect to volatility shocks, under the assumption that volatility follows a short-memory process. Christensen and Nielsen (2007) generalize their results to the case of a LM process, but do not consider the case of spurious LM. We consider the impact on the price elasticity of a spurious LM process for volatility.

Consider a stock price, P_t, satisfying

$$\frac{\mathbb{E}_t(P_{t+1}) - P_t + D_t}{P_t} = r_f + \alpha_t,$$

where r_f is the constant risk-free rate, D_t the dividend paid at time t and α_t the equity risk premium at time t, assumed to be linear in the variance

$$\alpha_t = \gamma\sigma_t^2.$$

Poterba and Summers (1986) show that if the variance follows a stationary AR(1) process

$$\sigma_{t+1}^2 - \bar{\sigma}^2 = \rho(\sigma_t^2 - \bar{\sigma}^2) + \mu_t,$$

then the elasticity of the stock market price with respect to volatility is

$$\frac{\mathrm{d}\log P_t}{\mathrm{d}\log\sigma_t^2} = -\frac{\bar{\alpha}}{1 + r_f + \bar{\alpha} - \rho(1+g)}, \tag{1}$$

where g is the constant growth rate of expected dividends and $\bar{\alpha}$ the mean risk premium.

Christensen and Nielsen (2007) derive the elasticity of the stock price with respect to volatility when the variance follows an ARFIMA process. They show that for such LM processes, the impact of volatility shocks on stock prices is small.

Let us consider instead a simple spurious LM process, i.e. an AR(1) process with a break in the coefficients. Let $y_t = \sigma_t^2 - \bar{\sigma}^2$. Then volatility evolves according to

$$y_{t+1} = \left[\rho_1 \mathbb{I}_{(t<T_B)} + \rho_2 \mathbb{I}_{(t\geq T_B)}\right] y_t + \mu_t, \tag{2}$$

where T_B is the break date and $\mathbb{I}$ is the indicator function.

The following proposition gives the elasticity of the stock price with respect to volatility for process (2)

Proposition 1. *Suppose the stock return volatility is governed by* (2)*, with break date* T_B*, then at time* $T_B - k$ *the elasticity of the stock level with respect to volatility, at the mean values of the risk premium and the dividend yield, is*

$$\frac{\mathrm{d}\log P_t}{\mathrm{d}\log \sigma_t^2} = \begin{cases} -\frac{\bar{\alpha}}{1+r_f+\bar{\alpha}-\rho_2(1+g)} & \textit{for } k = 0, -1, \dots \\ -\bar{\alpha}\left[\frac{1-\left(\frac{1+g}{1+r_f+\bar{\alpha}}\rho_1\right)^{k+1}}{1+r_f+\bar{\alpha}-\rho_1(1+g)} + \frac{\rho_2\rho_1^k\left(\frac{1+g}{1+r_f+\bar{\alpha}}\right)^{k+1}}{1+r_f+\bar{\alpha}-\rho_2(1+g)}\right] & \textit{for } k = 1, 2, \dots \end{cases} \tag{3}$$

Proof. From the chain rule it follows

$$\frac{\mathrm{d}P_t}{\mathrm{d}\sigma_t^2} = \mathbb{E}_t\left(\sum_{j=0}^{\infty} \frac{\mathrm{d}P_t}{\mathrm{d}\alpha_{t+j}} \frac{\mathrm{d}\alpha_{t+j}}{\mathrm{d}\alpha_t} \frac{\mathrm{d}\alpha_t}{\mathrm{d}\sigma_t^2}\right).$$

The first term inside the expectation is the derivative in an infinite order Taylor expansion of P_t around $\bar{\alpha}$, and is given by equation (6) of Poterba and Summers (1986) as

$$\frac{\mathrm{d}P_t}{\mathrm{d}\alpha_{t+j}} = -\frac{D_t(1+g)^j}{(1+r_f+\bar{\alpha})^{j+1}(r_f+\bar{\alpha}-g)}.$$

The second term is

$$\frac{\mathrm{d}\alpha_{t+j}}{\mathrm{d}\alpha_t} = \frac{\mathrm{d}y_{t+j}}{\mathrm{d}y_t} = \rho_1^j \mathbb{I}_{(t \le T_B - j)} + \sum_{h=1}^{j-1} \rho_1^{j-h} \rho_2^h \mathbb{I}_{(t = T_B-(j-h))} + \rho_2^j \mathbb{I}_{(t \ge T_B)}$$

and the third one

$$\frac{\mathrm{d}\alpha_t}{\mathrm{d}\sigma_t^2} = \gamma.$$

Setting $K_t = -\gamma D_t/[(1 + r_f + \bar{\alpha})(r_f + \bar{\alpha} - g)]$ and $c = (1+g)/(1+r_f+\bar{\alpha})$, this means that for $k = 1, 2, \ldots$

$$\sum_{j=0}^{\infty} \frac{\mathrm{d}P_t}{\mathrm{d}\alpha_{t+j}} \frac{\mathrm{d}\alpha_{t+j}}{\mathrm{d}\alpha_t} \frac{\mathrm{d}\alpha_t}{\mathrm{d}\sigma_t^2} = K_t \sum_{j=0}^{\infty} c^j \left[\rho_1^j \mathbb{I}_{(j \in \{0,\ldots,k\})} + \rho_1^k \rho_2^{j-k} \mathbb{I}_{(j \in \{k+1, k+2, \ldots\})}\right]$$

$$= K_t \left[\sum_{j=0}^{k} (c\rho_1)^j + \rho_1^k \rho_2 c^{k+1} \sum_{j=0}^{k} (c\rho_2)^j\right] = K_t \left[\frac{1-(c\rho_1)^{k+1}}{1-c\rho_1} + \frac{\rho_1^k \rho_2 c^{k+1}}{1-c\rho_2}\right].$$

For $k = 0, -1, \ldots$, instead

$$\sum_{j=0}^{\infty} \frac{\mathrm{d}P_t}{\mathrm{d}\alpha_{t+j}} \frac{\mathrm{d}\alpha_{t+j}}{\mathrm{d}\alpha_t} \frac{\mathrm{d}\alpha_t}{\mathrm{d}\sigma_t^2} = K_t \sum_{j=0}^{\infty} (c\rho_2)^j = \frac{K_t}{1-c\rho_2}.$$

For every k the desired result is obtained by noting that

$$\frac{\mathrm{d}\log P_t}{\mathrm{d}\log \sigma_t^2} = \frac{\sigma_t^2}{P_t} \frac{\mathrm{d}P_t}{\mathrm{d}\sigma_t^2}$$

and by setting the risk premium α_t and the dividend yield D_t/P_t equal to their mean values, $\bar{\alpha}$ and $r_f + \bar{\alpha} - g$ respectively. □

Contrary to the models of Poterba and Summers (1986) and Christensen and Nielsen (2007), under (2) the elasticity is not constant over time, but it depends on the distance between t and the break date T_B.

To evaluate the theoretical relevance of Proposition 1, we undertake a small numerical exercise. We calculate expression (3) by considering the same values for r_f, g and $\bar{\alpha}$ as in Christensen and Nielsen (2007). This allows us to measure the impact of k (i.e. how far behind is t from the break date T_B) and ρ_1 and ρ_2 on the elasticity. Note that for the above parameters, the elasticity estimates range from -0.006, when $d = 0$, to -0.028415 when $d = 0.49$ in an ARFIMA(0,d,0) for the volatility process (see Table 1 of Christensen and Nielsen, 2007, p. 688).

Table 1, Panel A, reports the stock price elasticity for alternative values of k. The first column refers to $\rho_1 = 0.6$, $\rho_2 = 0.6$ and $T_B = 90$. The second column refers to $\rho_1 = 0.7$, $\rho_2 = 0.99$, $T_B = 160$. In Panel B, we report the average LM parameter obtained simulating 1,000 time series according to (2), using the same sample size of the realized variance used in the original paper. It is worth noticing that when the break occurs in the middle of the sample, the elasticity is close to the one obtained by Christensen and Nielsen (2007). However, when the break occurs near the end of the sample the elasticity (in absolute value) we obtain is much larger. For a comparison, with $d = 0.3$ Christensen and Nielsen (2007) obtain an elasticity equal to -0.015.

[Table 1 about here.]

2.2. Cross-Sectional Aggregation and Spurious Long Memory

The analysis of this section is based on the aggregation of time series as in Granger (1980), Zaffaroni (2004), and Souza (2008). In particular, Zaffaroni (2004) studies the effect of the averaging of ARMA processes with a common and an idiosyncratic shock, with random coefficients. The author derives conditions that induce LM on the aggregate series. A similar aggregation exercise is proposed by Altissimo et al. (2009) and applied to CPI indexes. Their model explains the low persistence observed at the level of sectoral inflation and high persistence of the aggregate inflation. Genuine LM arises as the effect of aggregation also in the nonlinear model studied by Davidson and Sibbertsen (2005). In what follows, we study a simple case in which cross-sectional aggregation may generate spurious LM in the data.

Let us assume now that the series, y_{it} follows the process

$$y_{it} = \alpha_{it}\mathbb{I}_{(t<T_B)} + \beta_{it}\mathbb{I}_{(t\geq T_B)} + u_t + e_{it}, \qquad i = 1, \ldots, n \quad t = 1, \ldots, T \tag{4}$$

where the common shock u_t is an i.i.d. sequence $(0, \sigma_u^2)$, the idiosyncratic shock e_{it} is an i.i.d. sequence $(0, \sigma_i^2)$ for any i and t. For any i and t, the α_{it} (β_{it}) are i.i.d. with probability density f_α (f_β) over the support $(-1, 1)$.

It is possible to choose the distributions f_α and f_β so that

1. The n units do not show LM, but
2. The aggregate $Y_{n,t} = n^{-1}\sum_{i=1}^{n} y_{it}$ is a spurious (stationary) LM process

We perform a Monte Carlo experiment in which we simulate $M = 1,000$ times series y_{it} for each $i = 1, \ldots, n = 500$ and $t = 1, \ldots, T = 500$ according to model (4). The random variables α_{it} have a Beta distribution with parameters $a_1 = 2$ and $b_1 = 0.52$ whereas the random variables β_{it} are distributed according to a Beta distribution with parameters $a_2 = 4.5$ and $b_2 = 0.78$. Further, we chose $\sigma_u = 0.05$ and sampled the standard deviations σ_i independently from a uniform distribution on $(0.2, 0.4)$. For each Monte Carlo replication we build the aggregate time series $Y_{n,t}$ and estimated the LM parameter d using the local Whittle (LW) estimator for the aggregate and the n individual components. In addition, to detect the presence of spurious LM, we perform the test based on the profiled LW likelihood (LWL) with $\epsilon = 0.02$ on the $n + 1$ time series at our disposal. (See Appendix A, Section A.5 for details on the test.) In Table 2, we report the average value of the LM parameter for $Y_{n,t}$ (mean(d)) across the M replications, the average value of the mean LM parameter for the components (mean($\bar{d}$)) across the M replicas, the percentage of rejections of the null of true LM for the aggregate and the average percentage of rejections of the null of true LM for the single components.

[Table 2 about here.]

In the next section, we undertake a comprehensive Monte Carlo simulation exercise to compare the performance of five tests constructed under the null of true LM versus the alternative of spurious long memory due to level shifts or breaks. The tests are briefly reviewed in Appendix A.

3. TESTS AGAINST SPURIOUS LONG MEMORY: A MONTE CARLO COMPARISON

In this section, the size and the power of the five tests against spurious LM are evaluated under several alternatives, from a number of stochastic processes, e.g. the random level shift, the Stopbreak and the Markov-Switching model, to breaking processes. We evaluate the performance of the tests, allowing for a wide range of sizes and dates of the breaks. In this way not only we identify the best tests in terms of overall performance, we also describe their behavior under several features and show how each of them may modify the reliability of the tests. Thus we are able to suggest the best choices for the test given the underlying DGP.

Sections 3.1 and 3.2 report the empirical size and power of the five testing procedures, respectively. Throughout the paper, for the "Temporal Aggregation" test of Ohanissian et al. (2007) we set the number of aggregation levels to $N = 4$ and the aggregation levels to $n_j = 2^j$ for $j = 0, \dots, N-1$. For the "Sample Splitting" test of Shimotsu (2006), the number of subsamples considered is $N = 2$. We also consider $N = 4$ and $N = 8$ but we do not report the results given they are dominated by the $N = 2$ case. For the LWL test of Qu (2011), the trimming parameter is set equal to $\epsilon = 0.02$ or $\epsilon = 0.05$. For the GPH and the LW estimates, the bandwidth is $m = [T^{0.5}]$, where T denotes the sample size.

3.1. Evaluation of the empirical size of the tests

We simulate 10,000 series of dimension $T = 200, 600, 1000, 2000$ and 5000 using three DGPs:

1. ARFIMA(0,d,0) with $d = 0.4$;
2. ARFIMA(1,d,0) with $d = 0.4$, $\rho = 0.4$
3. ARFIMA(0,d,1) with $d = 0.4$, $\theta = 0.4$,

where ρ and θ denote the autoregressive and the moving average coefficient, respectively. Table 3 reports the rejection frequencies for the five tests using a 5% confidence level.

The "Temporal Aggregation" test has an empirical size which is very close to the theoretical 5% even in small samples. The "Sample Splitting" test is over sized whereas the SB-FDF and the LWL are much more conservative tests since they under-reject the null. The "Differencing" test shows a very low size in its PP version and slight under-rejection in the KPSS case. With respect to the sample size, the LWL test has a very low size in small sample and then it converges to the theoretical size when the sample size increases. The SB-FDF and the "Differencing" - KPSS case are characterized by a moderate increase in their size with larger samples whereas the "Sample Splitting" and the "Differencing" - PP case are almost constant across the different sample sizes. Finally, the presence of either an autoregressive or a moving average component does not alter the results.

[Table 3 about here.]

3.2. Evaluation of the empirical power of the tests

Section 3.2.1 is devoted to the analysis of size-adjusted powers when the underlying process is subject to a structural break, which spuriously induces LM. In Section 3.2.2, we evaluate the performance of the tests under several alternative break models. The size-adjustment is motivated by the fact that we are comparing tests with different behavior in terms of size, as shown in the previous section. In Appendix B, we report the size-unadjusted power of the tests, which shows very similar pattern of the size-adjusted one.

3.2.1 Case A: The impact of the location and the size of the break

In this section, we evaluate the size-adjusted power of the five tests under several structural break model alternatives that spuriously induce LM. To understand the size-adjusted power sensitivity to break size and timing of the break, we simulate 10000 series of dimension $T = 200, 600, 1000, 2000$ and 5000 using the data generating processes $Y_t = \alpha_2 DU_t(\eta) + u_t$. We consider different break intensities, i.e., a small ($\alpha_2 = 0.25$), a medium ($\alpha_2 = 0.5$) and a large break ($\alpha_2 = 0.75$), and different break times (i.e., break in the first quarter, in the middle and in the last quarter of the sample).

Table 4 reports the rejection frequencies for the five tests using a 5% confidence level. The size-adjusted power of the tests as a function of the number of observations in the time series is plotted in Figures 1–3.

[Table 4 and Figures 1–3 about here.]

The LWL test presents the highest rate of rejection, despite its weakness in detecting breaks in presence of small break size and sample size not sufficiently large. When the trimming parameter is set to $\epsilon = 0.05$, the test achieves a good power in small samples and poor results in larger ones, namely for $T \geq 1000$. Our results confirm Qu's suggestion to choose $\epsilon = 0.02$ as a good compromise between size and power. The second best is the SB-FDF test which detects both large and small breaks. The "Temporal Aggregation", the "Sample Splitting" and the "Differencing" tests show less power. For the former, the rate of rejection of the null is between 1% and 2% in small samples and between

5% and 30% when $T = 5000$. Such a poor performance is due to the fact that the number of aggregation levels ($N = 4$ in the experiment) considered is not big enough. However, increasing N results in aggregated series whose sample size is too limited to provide valid GPH estimates. The "Sample Splitting" has a low rate of rejection in small samples but its power improves in large samples, whereas the "Differencing" test in both its KPSS and Phillips-Perron versions gives very poor results, with a power almost equal to zero. Those results are a direct consequence of the choice of the spectral bandwidth. Indeed, the power of the tests increases when a larger number of frequencies is considered. Overall, the larger the break, the higher the fractional parameter and the better the tests power, the only exception being the "Sample Splitting" test. Note that the power of the "Sample Splitting" test decreases dramatically when the break is in the middle of the sample, that is, when the breaking and the splitting time coincide, since we used $N = 2$. In this way, the splitting cancels out the effects of the break and therefore the test can hardly detect the presence of a change in the series. Finally, the location of the break has insignificant effects on the power of the test.

3.2.2 Case B: The impact of alternative break models

To further understand the size-adjusted power of the five tests, we extend our analysis by considering models that generate spurious LM. The four alternative data generating processes are:

1. Nonstationary random level shift model (RLS-NS): $X_t = \mu_t + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0, 5)$, $\mu_t = \mu_{t-1} + v_t \eta_t$, $v_t \sim i.i.d.Binomial(1, 0.002)$, $\eta_t \sim i.i.d.N(0, 1)$.

2. Stationary random level shift model (RLS-S): $X_t = \mu_t + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0, 1)$, $\mu_t = (1 - v_t)\mu_{t-1} + v_t \eta_t$, $v_t \sim i.i.d.Binomial(1, 0.003)$, $\eta_t \sim i.i.d.N(0, 1)$.

3. STOPBREAK model (STOPBREAK): $X_t = \mu_t + \epsilon_t$, $\epsilon_t \sim i.i.d.N(0, 1)$, $\mu_t = \mu_{t-1} +$

$\frac{\epsilon_{t-1}^2}{180+\epsilon_{t-1}^2}\epsilon_{t-1}$.

4. Markov-Switching model (MS-IID): $X_t \sim i.i.d.N(-1, 6.253)$ if $s_t = 0$ and $X_t \sim i.i.d.N(1, 2.543)$ if $s_t = 1$, where s_t is a Markov process with transition probabilities $p_{01} = p_{10} = 0.02$.

We generate spurious LM processes via the above four alternative break models widely used in finance and macroeconomics, in particular to model realized volatility and inflation, respectively. The RLS-NS and RLS-S processes are used by Perron and Qu (2010) for modeling stock returns volatility. The values of the parameters chosen to generate the Binomial random variable in both the RLS-NS and RLS-S models imply infrequent breaks in the mean, a genuine feature of financial time series, with the RLS-NS and the RLS-S models having on average a level shift every 500 and 334 observations, respectively. Perron and Qu (2010) document that daily S&P 500 returns are characterized by level shifts every 535 observations. The STOPBREAK process of Engle and Smith (1999) generates time series subject to random structural shifts at random intervals. The authors show that this kind of process may describe the behavior of stock prices that move together for periods of time and then jump apart occasionally. We follow Shimotsu (2006) in choosing the parameters of this process to induce a spurious LM pattern. Under our parameterization, the conditional mean of the process is subject to rare changes that have permanent effects on it. Finally, for the MS-IID model, in our simulation exercise we use transition probabilities which at the same time accurately reflect empirical work and generate spurious LM. In particular, we use the Markov transition probabilities estimated in Evans and Wachtel (1993) where the near unity values of all diagonal elements of the transition matrix are needed to generate spurious LM, as documented by Diebold and Inoue (2001). The near unity property of all probabilities of permanence is not uncommon in modeling financial and economic time series. For instance, Cai (1994) fits a two-state switching-regime ARCH to monthly returns and estimates transition probabilities of 0.0122 and 0.0598. Hamilton and Susmel (1994) evaluate the transition matrix of a similar model with three states to study the volatility of weekly stock returns estimating diagonal elements that range from 0.9831 to 0.9924. Markov switching process have been applied to inflation as well, with changes in regimes that capture the different behavior of inflation expectations across different monetary regimes, as discussed in Benati (2008). Even for inflation, there is empirical evidence towards diagonal elements of the transition matrix that are near unity. For instance, Evans and Wachtel (1993) fit a two state MS-IID model to quarterly inflation and report a probability of remaining in either states close to 98 percent. Evans and Lewis (1995) give further evidence to this result estimating Markov transition probabilities of 0.064 and 0.039. Probabilities of permanence close to one imply rare and long-lasting shifts in inflation. This is in line with the results of Cogley and Sargent (2001, 2005), where the zero frequency component of the spectral density of inflation is accounted for most of the decline of inflation persistence during the last decades.

As in the previous exercise, we simulate 10000 series of dimension $T = 200, 600, 1000, 2000$ and 5000 for each of the alternative break models.

Table 5 reports the rejection frequencies for the five tests using a 5% confidence level. In Figure 4, we plot the tests size-adjusted power of the tests against T.

[Table 5 and Figure 4 about here.]

The best performance is achieved by the two LWL tests, with an average power converging to 0.9 in large samples. The test with the trimming parameter $\epsilon = 0.02$ dominates the $\epsilon = 0.05$-based test, even if the latter achieves better results in small samples, as already noted in Section 3.2.1. The second best is the SB-FDF test, with results similar to the LWL in the RLS-NS and in the Stopbreak model. Yet, the test continues to show poor power when the Markov Switching model generates spurious LM.

The "Temporal Aggregation" test has rejection frequencies ranging between 7% and 30% for $T = 2000$, and between 35% and 83% for $T = 5000$. For the "Sample Splitting" test the rejection frequencies are between 10% and 20% when $T = 2000$ and between 10% and 43% when $T = 5000$. Overall, the "Differencing" test has a low power, and the KPSS version of this test achieves a larger power than the Phillips-Perron in all models but the Markov Switching.

To better understand the poor performance of the "Sample Splitting" and the "Differencing" tests, we replicate the Monte Carlo analysis allowing for two different choices of spectral bandwidth, namely $m = \left[T^{0.7}\right]$ and $m = \left[T^{0.9}\right]$. The choice of these two bandwidths follows from the Monte Carlo studies of Shimotsu (2006) and aims to show test reliability only at such high frequencies. We use the same structural break processes of Section 3.2.1 with the same sample size of the previous experiment. The results, not reported here but are available from the authors upon request, show reliable performance of the tests only when large spectral bands are considered. Furthermore, the $N = 2$ version of the "Sample Splitting" shows better performance than the $N = 4$ and $N = 8$ cases, whereas all three versions give a poor power whenever the date of the break coincides with the splitting date. The "Differencing" test in its KPSS version provides very good results whereas the Phillips-Perron version gives a zero power. This result, together with the near to unity power of the KPSS version, implies that the "Differencing" test wrongly detects the structural break processes as $I(1)$ processes. Thus, the very strong power is obtained with a over-differencing procedure, due to a bias in the fractional parameter estimates rooted in the choice of the spectral bandwidth, as well noticed by Robinson (1994, 1995).

4. CONCLUSIONS

Long memory in financial time series could arise as a spurious effect due to omitted level shifts or structural breaks. This paper has two main contributions. First, we developed some theoretical results on how a spurious LM volatility process affects the elasticity of the stock market price with respect to volatility; we also showed that spurious persistence in the data can be the effect of cross-sectional aggregation of units exhibiting a common and an idiosyncratic shock. The second main contribution is the evaluation of the performance of the five available tests against spurious LM. We studied their size and power via an extensive Monte Carlo experiment, that accommodates a wide range of possible alternatives, from genuine LM processes to series subject to breaks with different sizes and dates and a number of alternative stochastic processes, e.g. the random level shift , the Stopbreak and the Markov-switching model. Overall the best performances are given by the LWL test proposed by Qu (2011), with the SB-FDF of Dolado et al. (2005) as second best.

Our theoretical results highlight the important consequence of a spurious LM volatility and how spurious LM may arise when aggregating data. In addition, from the simulation study some useful guidelines for empirical work can be drawn. All tests achieved a suitable power function with sample size $T \geq 1000$. Thus, whereas we can apply the tests in financial time series, their application in economic time series with small samples needs further studies. Moreover, the larger the break, the higher the power, whereas the date of the break has almost no influence on the power of the tests. Even if occasional large breaks induce LM, the presence of small breaks has to be dealt with a lot of care. We suggest the joint implementation of the LWL and the SB-FDF tests, to achieve a suitable compromise between power against pure and alternative breaking models.

The findings in this paper suggest some further developments. For instance it is interesting to compare the asymptotic efficiency of the available test on spurious LM and compute the analytical bias in the fractional parameter as a function of the number, the size and the date of breaks. Further, since many economic and financial time series are non-stationary, an extension to such cases maybe useful for practical purposes. We leave these developments to future work.

APPENDICES

A. TESTING FOR TRUE VS SPURIOUS LONG MEMORY

In this section, we briefly review the five existing procedures to test for true versus spurious LM in stationary processes[1]. We refer to the original version of the papers for a more detailed description of the tests.

A.1. A Test Based on Temporal Aggregation

A direct implication of the scaling and self-similarity properties of the fractional Brownian motion (Mandelbrot and Van Ness, 1968) is that the stochastic process presents the same memory at all levels of sampling frequency. A formal test procedure which exploits the invariance of the LM parameter for temporal aggregates of the process under the null of true LM is proposed by Ohanissian et al. (2007). The test is applicable to a

[1]It is worth noticing that a number of tests, constructed under the null hypothesis of occasional breaks, are available in the literature. Bisaglia and Gerolimetto (2009) propose a test procedures based on the Box-Pierce and Ljung-Box statistics, while Berkes et al. (2006) construct a test based on the CUSUM statistic. Mayoral (2012) develops a time-domain test for non-stationary processes, under the null of $I(d)$ versus $I(0)$ plus trends and/or breaks. Rea et al. (2009) study the performance of the atheoretical regression trees procedure to identify breaks in LM processes. Finally, Kapetanios and Shin (2011) propose a Wald test under the null of nonstationary LM against the alternative of exponential smooth transition autoregressive processes.

stationary mean zero Gaussian LM time series $\{Y_t\}$. The spectral density is assumed to be $f(\lambda) = |1 - \mathrm{e}^{-\mathrm{i}\lambda}|^{-2d} f^*(\lambda)$, $0 < d < 0.5$, where $f^*(\cdot)$ denotes the spectral density of the short memory component of $\{Y_t\}_{t=1,\ldots,T}$ and is assumed to be continuous, bounded from above, bounded away from zero, twice differentiable with the second derivative bounded in a neighborhood of zero. This assumption allows the use of the GPH estimator to avoid many potential mis-specification issues involving the short memory component. Let n and T/n be finite positive integers. The n-period non-overlapping aggregates of the time series Y_t are defined as

$$Z_s^{(n)} = \sum_{\tau=1}^{n} Y_{n(s-1)+\tau} \quad \text{for } s = 1, \ldots, \frac{T}{n}.$$

With the convention that a superscript (n) on any statistic represents the corresponding statistic for the n-temporally aggregated data, let

$$I_j^{(n)} = \frac{n}{2\pi T} \left| \sum_{s=1}^{T} Y_s \exp\left(\left[\frac{s}{n}\right] \mathrm{i}\lambda_j n \right) \right|^2 = \frac{n}{2\pi T} \sum_{s=1}^{T} \sum_{t=1}^{T} Y_s Y_t \cos\left(\lambda_j n \left(\left[\frac{s}{n}\right] - \left[\frac{t}{n}\right] \right) \right)$$

where $\lambda_j = \frac{2\pi j}{T}$, for $j = 1, ..., \frac{T}{n}$, and $[\cdot]$ denotes the integer part operator. The periodogram of the n-temporally aggregated series, $I^{(n)}$ can be expressed in terms of the periodogram of the entire series, $I_j^{(1)}$:

$$I_j^{(n)} = n I_j^{(1)} + Y' B_j Y$$

where $Y = (Y_1, \ldots, Y_T)'$, $B_j = [b_j(s,t)]_{1 \le s,t \le T}$ and

$$b_j(s,t) = \frac{n}{2\pi T} \left(\cos\left(\lambda_j n \left(\left[\frac{s}{n}\right] - \left[\frac{t}{n}\right] \right) \right) - \cos(\lambda_j n) \right).$$

Given the bandwidth parameter $m^{(n)}$, the GPH estimate of the LM parameter for the n-temporally aggregated data is

$$\hat{d}^{(n)} = -\frac{1}{2\sum_{j=1}^{m^{(n)}} \left(a_j^{(n)}\right)^2} \sum_{j=1}^{m^{(n)}} a_j^{(n)} \log I_j^{(n)},$$

with

$$a_j^{(n)} = \log\left|\sin\left(\lambda_j n\right)\right| - \frac{1}{m^{(n)}} \sum_{j=1}^{m^{(n)}} \log\left|\sin\left(\lambda_j n\right)\right|.$$

The number of ordinates of the GPH estimator depends on the length of the series and since temporal aggregation decreases the length of the series, we have $m^{(n_1)} > m^{(n_2)}$ for $n_1 < n_2$ and also the frequencies are not the same: $\lambda_j^{(n_1)} \neq \lambda_j^{(n_2)}$ for $n_1 \neq n_2$ and $\lambda_j^{(kn)} = \lambda_{kj}^{(n)}$.

Let N denote the fixed, but arbitrarily large, number of aggregation levels and $(n_1, n_2, \ldots, n_N)$ the fixed, but arbitrarily large, aggregation levels for the N series such that $n_1 < n_2 < \ldots < n_N$. Ohanissian et al. (2007) show that, if the growth rate of the ordinates of the GPH estimator is such that $m^{(n)} = o(T^{\frac{2-4d}{3}})$ as $T \to \infty$, for any aggregation level n, then the joint distribution of the GPH estimates of the aggregated series is asymptotically normal. Furthermore, authors prove that, asymptotically, the covariance between any two GPH estimates obtained using temporally aggregated series equals the variance of the lesser aggregated series:

$$4m^{(n_i)}\left(\mathrm{Cov}(\hat{d}^{(n_i)}, \hat{d}^{(n_j)}) - \mathrm{Var}(\hat{d}^{(n_i)})\right) = o(1) \quad \text{as } T \to \infty \quad \text{for } 1 \leq i < j \leq N. \tag{5}$$

Equation (5) allows us to compute the theoretical covariance matrix. When the time series of interest is relatively small, the approximation suggested by GPH should be used instead:

$$\text{Var}_{\text{approx.}}(\hat{d}^{(n_i)} - d) = \left[24 \sum_{j=1}^{m^{(n_i)}} \left(a_j^{(n_i)} \right)^2 \right]^{-1} \pi^2. \tag{6}$$

Let $\widehat{\mathbf{d}}_N = (\hat{d}^{(n_1)}, \hat{d}^{(n_2)}, \ldots, \hat{d}^{(n_N)})'$ be the N-dimensional vector of the estimated LM parameters, and $\mathbf{d}_N = (d^{(n_1)}, d^{(n_2)}, \ldots, d^{(n_N)})'$ be the constant N-dimensional vector of the actual LM parameters. The null hypothesis is

$$\text{H}_0 : d^{(n_1)} = d^{(n_2)} = \ldots = d^{(n_N)} = d$$

and it can be tested by considering the quadratic form

$$W = \left(\widehat{\mathbf{d}}_N - \mathbf{d}_N \right)' \Lambda^{-1} \left(\widehat{\mathbf{d}}_N - \mathbf{d}_N \right) \sim \chi^2(N)$$

where Λ is the asymptotic covariance matrix that follows from (6). Note that for the asymptotic covariance matrix to be invertible the asymptotic variances of each individual GPH estimate must differ. This can be achieved by using a different number of ordinates for the estimation of each temporally aggregated series. Since in practice $\mathbf{d}$ is unknown the mean value of the estimates is used and the test statistic to implement is

$$\widehat{W}_N = (P\widehat{\mathbf{d}}_N)'(P\Lambda P')^{-1}(P\widehat{\mathbf{d}}_N) \sim \chi^2(N-1)$$

where

$$P = \begin{pmatrix} 1 - \frac{1}{N} & -\frac{1}{N} & -\frac{1}{N} & \cdots & -\frac{1}{N} & -\frac{1}{N} \\ -\frac{1}{N} & 1 - \frac{1}{N} & -\frac{1}{N} & \cdots & -\frac{1}{N} & -\frac{1}{N} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ -\frac{1}{N} & -\frac{1}{N} & -\frac{1}{N} & \cdots & 1 - \frac{1}{N} & -\frac{1}{N} \end{pmatrix}.$$

Since $P\widehat{\mathbf{d}}$ is $N(0, P\Lambda P')$, the test statistic $\widehat{W_N}$ has an asymptotic $\chi^2(N-1)$ distribution under H_0.

A.2. The Structural Break–Fractional Dickey–Fuller Test

The Structural Break–Fractional Dickey–Fuller (SB–FDF) test is developed in Dolado et al. (2005). Under the null, the process of interest is $I(d)$, $d \in (0,1)$. Under the alternative, the process is $I(0)$ with one structural break. To account for structural breaks, the process is assumed to be described by the following equation:

$$Y_t = A_B(t) + \frac{a_t \mathbb{I}_{(t>0)}}{\Delta^d - \phi L} \tag{7}$$

where $A_B(t)$ is a linear deterministic trend function that may contain breaks at the unknown date T_B, a_t is a stationary $I(0)$ process, L the lag operator $\Delta = 1 - L$, and $\mathbb{I}$ the indicator function. Under the null $H_0 : \phi = 0$, whereas $\phi < 0$ means that the process is $I(0)$ and it is subject to the regime shifts defined by $A_B(t)$. Rearranging (7) yields:

$$\Delta^d Y_t = \phi Y_{t-1} + \Delta^d A_B(t) - \phi A_B(t-1) + \varepsilon_t \tag{8}$$

where $\varepsilon_t = a_t \mathbb{I}_{(t>0)}$. The SB–FDF test of $I(d)$ vs. $I(0)$ in the presence of structural breaks is simply given by the t-ratio of the coefficient ϕ in the regression model (8). The definition of $A_B(t)$ we consider is $A_B(t) = \mu_0 + (\mu_1 - \mu_0)DU_t(\eta)$ which corresponds to the so-called *crash* hypothesis. Here $DU_t(\eta) := \mathbb{I}_{(t>T_B)}$, where T_B, usually expressed as a fraction of the sample size, $T_B = \eta T$, is the date of the break. Dolado et al. (2005) prove that for a process generated by (7) with $a_t \sim i.i.d(0, \sigma^2)$ the asymptotic distribution of the test statistic under the null of $\phi = 0$, when ϕ is estimated by OLS is non-standard if $d \in (0.5, 1)$ and standard normal if $d \in (0, 0.5)$.

When, more realistically, the break date T_B is unknown, one may choose the break point that gives the least favorable result for the null hypothesis of $I(d)$ using the SB-

FDF test in (8). Therefore, following Andrews (1993), the t-statistic of the coefficient ϕ is computed for the values of $\eta \in (0.15, 0.85)$ and the most negative value is chosen:

$$\widehat{t}_{\hat{\phi}} = \inf_{\eta \in (0.15, 0.85)} t_{\hat{\phi}(\eta)}.$$

The choice of $\eta \in (0.15, 0.85)$ guarantees the uniform consistency of the test statistics and rules failures of the continuous mapping theorem out. Dolado et al. (2005) show that under the null and when $\mu_0 = \mu_1$, the asymptotic distribution of $\widehat{t}_{\hat{\phi}}$ is again non-standard if $d \in (0.5, 1)$ and standard normal if $d \in (0, 0.5)$. In the former case, the simulated critical values are provided by Dolado et al. (2005, Appendix B). The test reject the null of genuine LM when $\widehat{t}_{\hat{\phi}} < k_{\inf,\alpha}$, where $k_{\inf,\alpha}$ is the simulated critical value at the significance level α.

A.3. A Test Based on Sample Splitting

The test proposed by Shimotsu (2006) is constructed by splitting the sample into b subsamples, estimating d for each subsample and computing how these estimates differ from the full sample estimate. Under the null of true LM each subsample follows an $I(d)$ process with the same value of the LM parameter d.

Let N and T/N be finite positive integers. The sample $\{Y_s\}_{s=1,\dots,T}$ is split into N blocks, so that each block has T/N observations. The author suggests to compute the local Whittle (LW) estimator (Robinson, 1995) for each block of observations using m/N periodogram ordinates, m being the number of periodogram ordinates used for the estimation of d in the full sample. Let $\tilde{d}$ be the LW estimator of $d \in (-0.5, 0.5)$ computed from the whole sample and $\tilde{d}^{(n_i)}$, $(i = 1, 2, \dots, N)$, be the LW estimator of d computed from the ith block of the observations, $\{Y_t : t = (n_i - 1)T/N + 1, \ldots, n_i T/N\}$:

$$\tilde{d}^{(n_i)} = \arg \min_{-\frac{1}{2}<d<\frac{1}{2}} \left\{ \log \left[\frac{N}{m} \sum_{j=1}^{m/N} \tilde{\lambda}_j^{2d} I_j^{(n_i)} \right] - 2d \frac{N}{m} \sum_{j=1}^{m/N} \log \tilde{\lambda}_j \right\} \quad i = 1, \ldots, N$$

$$I_j^{(n_i)} = \frac{N}{2\pi T} \left| \sum_{s=(n_i-1)T/N+1}^{n_i T/N} Y_s \exp\left(\mathrm{i} s \tilde{\lambda}_j\right) \right|^2,$$

where $\tilde{\lambda}_j = \frac{2\pi j N}{T}$, for $j = 1, \ldots, T/N$. The test statistic under the null of genuine long memory is then given by

$$\widetilde{W}_N = (A\widehat{\mathbf{d}}_N)'(A\Upsilon A')^+(A\widehat{\mathbf{d}}_N)$$

where $\widetilde{\mathbf{d}}_N$ is a the $N+1$ column vector

$$\widetilde{\mathbf{d}}_N = \begin{pmatrix} \tilde{d} - d & \tilde{d}^{(1)} - d & \cdots & \tilde{d}^{(N)} - d \end{pmatrix}',$$

$\Upsilon = \begin{pmatrix} 1 & \iota_N' \\ \iota_N & I_N \end{pmatrix}$ is the covariance matrix of $\widetilde{\mathbf{d}}_N$ and $A = [\ \iota_N \ | - I_N]$. Here ι_N denotes a $N \times 1$ vector of ones, I_N is the identity matrix of dimension N and P^+ denotes the pseudo-inverse of the matrix P. Note also that $A\Upsilon A' = NI_N - \iota_N \iota_N'$. The test statistic $\widetilde{W}_N$ has a chi-squared limiting distribution with $N-1$ degrees of freedom. In case of non-stationary LM, the test is based on the version of the LW estimator introduced in Shimotsu and Phillips (2005).

A.4. Test using dth Differencing

An alternative test proposed by Shimotsu (2006) is based on the fact that if an $I(d)$ process is differenced d times, then the resulting time series is trivially an $I(0)$ process.

The test is constructed by demeaning the data and applying the Phillips-Perron (PP) unit root test or the KPSS test to its $\hat{d}$th difference, $\hat{d}$ being a consistent estimate of d.

Some care is needed in demeaning the data. Assuming that Y_t follows a truncated $I(d)$ process with initialization at $t = 0$:

$$Y_t - \mu = (1 - L)^{-d} u_t \mathbb{I}_{(t \geq 1)}$$

where μ denotes either the mean of Y_t in case $d < 1/2$ or the initial condition of Y_t in case $d > 1/2$, Shimotsu (2006) shows that

$$(1 - L)^{-d} \left(Y_t - T^{-1} \sum_{t=1}^{T} Y_t \right) = u_t + \mathrm{O_P}(T^{d-1/2} t^{-d}) \tag{9}$$

given that $T^{-1} \sum_{t=1}^{T} Y_t - \mu = \mathrm{O_P}(T^{d-1/2})$. Hence, if $d \geq 1$, the second term on the right of (9) has a nonnegligible effect on the sample statistics of the dth differenced demeaned data.

To circumvent the problem, the author suggests to estimate μ using a linear combination of the sample mean and Y_1:

$$\hat{\mu}(d) = \frac{w(d)}{T} \sum_{t=1}^{T} Y_t + (1 - w(d)) Y_1 \tag{10}$$

where $w(d)$ is a is a smooth weight function such that $w(d) = 1$ for $d \leq 1/2$ and $w(d) = 0$ for $d \geq 3/4$. For suitable choices of the weight function, we refer to Shimotsu (2006).

Once the data has been demeaned using (10) to estimate the mean, it is possible to compute the $\hat{d}$th differenced series U_t:

$$U_t = (1 - L)^{\hat{d}} (Y_t - \hat{\mu}(\hat{d})) = \sum_{k=0}^{t-1} \frac{\Gamma(k - \hat{d})}{\Gamma(-\hat{d}) k!} (Y_{t-k} - \hat{\mu}(\hat{d})).$$

Finally, a stationarity test (PP or KPSS) is applied to the series U_t.

A.5. A Test Based on The Local Whittle Likelihood

Qu (2011) proposes a test based on the profiled LW likelihood (LWL) function that exploits the different behavior of true and spurious LM processes around some critical frequencies. The mean of the spectral density of genuine long processes is proportional to λ_j^2 uniformly for $\lambda_j = o(1)$ whereas, in short memory processes, level shifts affect the periodogram around a neighborhood of $\lambda_j = O(T^{-1/2})$ frequencies, as shown by Perron and Qu (2010). Generalizing this result, the author introduces a test based on the profiled likelihood function of the LW estimator, within the Kolmogorov-Smirnov type test. The statistic is based on a process of weighted periodograms to exploit the behavior of spurious and genuine LM processes in their critical ranges, i.e. around $\lambda_j = T^{-1/2}$. Consider the LW likelihood function

$$Q\left(G,d\right) = \frac{1}{T}\sum_{j=1}^{m}\left\{\log G\lambda_j^{-2d} + \frac{I_j}{G\lambda_j^{-2d}}\right\}, \tag{11}$$

where $\lambda_j = \frac{2\pi j}{T}$, for $j = 1, ..., \frac{T}{n}$. The profiled likelihood function is derived by minimizing (11) with respect to G:

$$R\left(d\right) = \log G\left(d\right) - 2d\frac{1}{m}\sum_{j=1}^{m}\log\lambda_j,$$

where

$$G\left(d\right) = \frac{1}{m}\sum_{j=1}^{m}\lambda_j^{2d}I_j.$$

The test statistics is defined as

$$\widetilde{\widetilde{W}}_\epsilon = \sup_{r\in[\epsilon,1]}\left(\sum_{j=1}^{m}v_j^2\right)^{-1/2}\left|\sum_{j=1}^{[mr]}v_j\left(\frac{I_j}{G(\hat{d})\lambda_j^{-2\hat{d}}} - 1\right)\right|,$$

where

$$v_j = \log \lambda_j - \frac{1}{m} \sum_{j=1}^{m} \log \lambda_j,$$

$\epsilon \in [0, 1]$ is a trimming parameter and $\hat{d}$ is the LW estimate of the fractional parameter computed with m frequency components. The presence of the trimming parameter ϵ ensures a reliable asymptotic approximation even in small samples. The author suggests to choose $\epsilon = 0.02$, though $\epsilon = 0.05$ performs well in small samples.

B. SIZE UNADJUSTED POWER OF THE TESTS

Table 6 reports the rejection frequencies in the case of a structural break model for the five tests using a 5% confidence level. Table 7 reports the rejection frequencies for the five tests using a 5% confidence level.

[Table 6 and 7 about here.]

ACKNOWLEDGMENTS

We are grateful to participants in the New York Camp Econometrics IV (The Mirror Inn, Lake Placid, NY, USA, 3-5 April 2009) in particular Badi Baltagi, Chihwa Kao and Hashem Pesaran; the Econometric Society 2009 North American Summer Meeting (Boston University, USA, 4-7 June 2009) in particular Jean-Marie Dufour, Domenico Giannone, Zhongjun Qu; the 15th International Conference Computing in Economics and Finance (University of Technology, Sydney, Australia, 15-17 July 2009); the 2009 Far East and South Asia Meeting of the Econometric Society (University of Tokyo, Faculty of Economics, Tokyo, Japan, 3-5 August 2009); the London-Oxbridge Time Series Workshop (Centre for International Macroeconomics and Finance, University of Cambridge, UK, 9 October 2009) in particular Andrew Harvey and Hashem Pesaran, for useful discussions and suggestions. Useful comments were provided from a lengthy discussion with Juan Dolado. We thank Morten Nielsen for useful comments especially on Section 2. We wish to thank the Editor, Esfandiar Maasoumi, an Associate Editor and an anonymous referee for useful comments and suggestions. However, the usual disclaimer applies. We acknowledge financial support from the Centre for Econometric Analysis (CEA@Cass), the 2008 and 2012 Cass Pump-Priming grant, and the 2008 City Pump-Priming grant and from the Department of Economics "Hyman P. Minsky" of University of Bergamo (2009 Fondi Ricerche di Ateneo ex 60%).

REFERENCES

Altissimo, F., Mojon, B., Zaffaroni, P. (2009). Can aggregation explain the persistence of inflation? *Journal of Monetary Economics* 56: 231–241.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2001). The distribution of realized exchanged rate volatility. *Journal of the American Stastistical Association* 96: 42–55.

Andersen, T. G., Bollerslev, T., Diebold, F. X., Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica* 71: 579–625.

Andrews, D. W. (1993). Test for parameter instability and structural break with unknown change point. *Econometrica* 61: 821–856.

Baillie, R. T., Kapetanios, G. (2007). Testing for neglected nonlinearity in long-memory models. *Journal of Business and Economic Statistics* 25: 447–461.

Benati, L. (2008). Investigating inflation persistence across monetary regimes. *The Quarterly Journal of Economics* 123: 1005–1060.

Berkes, I., Horváth, L., Kokoszka, P., Shao, Q.-M. (2006). On discriminating between long-range dependence and changes in mean. *The Annals of Statistics* 34: 1140–1165.

Bisaglia, L., Gerolimetto, M. (2009), Testing structural breaks vs. long memory with the Box-Pierce statistics: a Monte Carlo study. *Statistical Methods and Applications* 18: 543-553.

Cai, J. (1994). A Markov model of switching-regime ARCH. *Journal of Business & Economic Statistics* 12: 309–316.

Choi, K., Zivot, E. (2007). Long memory and structural changes in the forward discount: An empirical investigation. *Journal of International Money and Finance* 26: 342–363.

Christensen, B. J., Nielsen, M. Ø. (2007). The effect of long memory in volatility on stock market fluctuations. *Review of Economics and Statistics* 89: 684-700.

Cogley, T., Sargent, T. (2001). Evolving post-world war II u.s. inflation dynamics. *NBER Macroeconomics Annual* 16: 331–373.

Cogley, T., Sargent, T. (2005). Drift and volatilities: Monetary policies and outcomes in the post WWII U.S. *Review of Economic Dynamics* 8: 262–302.

Davidson, J., Sibbertsen, P. (2005). Generating schemes for long memory processes: regimes, aggregation and linearity. *Journal of Econometrics* 128: 253–282.

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

Dolado, J. J., Gonzalo, J., Mayoral, L. (2005). What is what?: A simple time-domain test of long-memory vs. structural breaks. Working Paper 954, UPF, Department of Economics.

Engle, R. F., Smith, A. D. (1999). Stochastic permanent breaks. *The Review of Economics and Statistics* 81: 553–574.

Evans, M., Lewis, K. K. (1995). Do expected shifts in inflation affect estimates of the long-run Fisher relation? *Journal of Finance* 50: 225–253.

Evans, M., Wachtel, P. (1993). Inflation regimes and the sources of inflation uncertainty. *Journal of Money, Credit and Banking* 25: 475–511.

Geweke, J., Porter-Hudak, S. (1983). The estimation and application of long memory time series models. *Journal of Time Series Analysis* 4: 221–238.

Granger, C. W. J. (1980). Long memory relationships and the aggregation of dynamic models. *Journal of Econometrics* 14: 227–238.

Granger, C. W. J., 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.

Hamilton, J. D., Susmel, R. (1994). Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics* 64: 307–333.

Hyung, N., Poon, S.-H., Granger, C. W. J. (2007). The source of long memory in financial market volatility. In Rapach, D. E., Wohar, M. E. (Eds.), *Forecasting in the Presence of Structural Breaks and Model Uncertainty*. Elsevier Series.

Kapetanios, G., Shin, Y. (2011). Testing the null hypothesis of nonstationary long memory against the alternative hypothesis of a nonlinear ergodic model. *Econometric Reviews* 30: 583–608.

Mandelbrot, B. B., Van Ness, J. W. (1968). Fractional Brownian motions, fractional noises and applications. *SIAM Review* 10: 422-437.

Mayoral, L. (2012). Testing for fractional integration versus short memory with trends and structural breaks. *The Oxford Bulletin of Economics and Statistics* 74: 278–305.

McAleer, M., Medeiros, M. C. (2008). A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries. *Journal of Econometrics* 147: 104–119.

Ohanissian, A., Russell, J. R., Tsay, R. S. (2008). True or spurious long memory? A new test. *Journal of Business & Economic Statistics* 26: 161–175.

Perron, P., Qu, Z. (2010). Long memory and level shifts in the volatility of stock market return indices. *Journal of Business & Economic Statistics* 28: 275–290.

Poterba, J. M. and L. H. Summers (1986). The persistence of volatility and stock market fluctuations. *The American Economic Review 76*(5), 1142–1151.

Qu, Z. (2011). A test against spurious long memory. *Journal of Business & Economic Statistics 29*(3), 423–438.

Rea, W. S., Reale, M., Cappelli, C., Brown, J. A. (2009). Identification of changes in mean with regression trees: An application to market research. *Econometric Reviews* 29: 754–777.

Robinson, P. M. (1994). Rates of convergence and optimal spectral bandwidth for long range dependence. *The Annals of Statistics* 99: 443–473.

Robinson, P. M. (1995). Gaussian semiparametric estimation of long range dependence. *The Annals of Statistics* 23: 1630–1661.

Shimotsu, K. (2006). Simple (but effective) tests of long memory versus structural breaks. Working Paper 1101, Queen's University.

Shimotsu, K., Phillips, P. C. (2005). Exact local whittle estimation of fractional integration. *The Annals of Statistics* 33: 1890–1933.

Smith, A. (2005). Level shifts and the illusion of long memory in economic time series. *Journal of Business & Economic Statistics* 23: 321–335.

Souza, L. R. (2008). Why aggregate long memory time series? *Econometric Reviews* 27: 298–316.

Zaffaroni, P. (2004). Contemporaneous aggregation of linear dynamic models in large economies. *Journal of Econometrics 120*(1), 75–102.

TABLE 1: Stock price elasticity using formula (3) for parameters $r_f = 0.035$, $g = 0.00087$ and $\bar{\alpha} = 0.006$. In the bottom panel the average d estimate from simulating 1,000 time series of length $T = 180$ for y_t using (2) are reported. The d estimates are based on a bandwidth $\sqrt{T}$.

Panel A: Elasticity		
k	$\rho_1 = 0.6$, $\rho_2 = 0.6$, $T_B = 90$	$\rho_1 = 0.7$, $\rho_2 = 0.99$, $T_B = 160$
≤ 0	-0.0315	-0.1197
1	-0.0240	-0.0863
2	-0.0196	-0.0638
3	-0.0171	-0.0487
4	-0.0156	-0.0386
5	-0.0148	-0.0317
Panel B: Long Memory Estimates		
$\bar{d}$ LW	0.2671	0.2834
$\bar{d}$ ELW	0.2727	0.3158
$\bar{d}$ FELW	0.2729	0.3197

TABLE 2: Long memory parameter estimates and LWL tests against spurious long memory for the aggregate and the components. Results are based on $M = 1,000$ Monte Carlo replicas. 'mean(d)' and 'mean($\bar{d}$)' are the average values across the M replicas of the long memory parameter for the aggregate and of the average long memory parameter for the components. The percentage of rejections of the null of true long memory for the aggregate and the the components are also reported.

mean(d)	0.2765
mean($\bar{d}$)	-0.0533
Rejections of H_0 for $Y_{n,t}$ (%)	80.50%
Average Rejection of H_0 for the components (%)	0.42%

TABLE 3: Rejection frequencies for ARFIMA models. The rejection frequencies for an ARFIMA(0,d,0), an ARFIMA(1,d,0) and an ARFIMA(0,d,1), with $d = 0.4$ and sample sizes $T = 200, 600, 1000, 2000, 5000$ are reported. The autoregressive and the moving average coefficient are both 0.4. Theoretical size is 5%.

	mean$(\hat{d})$	Temp. Aggreg.	SB–FDF	Sample Split.	Diff. PP	Diff. KPSS	LWL $\epsilon = 0.02$	LWL $\epsilon = 0.05$
				ARFIMA(0,d,0), $d = 0.4$.				
$T = 200$	0.3811	0.0461	0.0212	0.0913	0.0062	0.0300	0.0087	0.0096
$T = 600$	0.3950	0.0562	0.0256	0.0864	0.0071	0.0329	0.0113	0.0137
$T = 1000$	0.3929	0.0524	0.0294	0.0835	0.0097	0.0366	0.0181	0.0239
$T = 2000$	0.3960	0.0521	0.0328	0.0771	0.0121	0.0381	0.0248	0.0284
$T = 5000$	0.3977	0.0495	0.0411	0.0758	0.0218	0.0398	0.0455	0.0463
				ARFIMA(1,d,0), $d = \rho = 0.4$.				
$T = 200$	0.3926	0.0565	0.0165	0.0822	0.0012	0.0403	0.0097	0.0135
$T = 600$	0.4215	0.0572	0.0200	0.0813	0.0023	0.0389	0.0101	0.0164
$T = 1000$	0.4172	0.0567	0.0232	0.0784	0.0093	0.0421	0.0143	0.0188
$T = 2000$	0.4115	0.0573	0.0277	0.0759	0.0126	0.0409	0.0212	0.0257
$T = 5000$	0.3742	0.0571	0.0389	0.0702	0.0245	0.0414	0.0336	0.0417
				ARFIMA(0,d,1), $d = \theta = 0.4$.				
$T = 200$	0.4056	0.0642	0.0201	0.0856	0.0137	0.0312	0.0091	0.0107
$T = 600$	0.4201	0.0634	0.0236	0.0831	0.0143	0.0328	0.0103	0.0126
$T = 1000$	0.4254	0.0611	0.0284	0.0816	0.0268	0.0354	0.0157	0.0194
$T = 2000$	0.4166	0.0586	0.0308	0.0793	0.0301	0.0376	0.0187	0.0259
$T = 5000$	0.3752	0.0579	0.0376	0.0771	0.0355	0.0392	0.0274	0.0314

NOTE: Temp. Aggreg. is the "Temporal Aggregation" test with $N = 4$ aggregation levels. SB-FDF is the "Structural Break-Fractional Dickey-Fuller Test". Sample Split. is the "Sample Splitting" test with $b = 2$ subsamples. Diff. KPSS denotes the KPSS d-th Differencing test whereas Diff. PP is its Phillips-Perron version. LWL denotes the test based on the local Whittle likelihood. For each test the null is true long memory. For the GPH and LW estimators the spectral bandwidth is $m = \left[\sqrt{T}\right]$. Results are based on 10000 Monte Carlo replications.

TABLE 4: Size-adjusted rejection frequencies at a 5% confidence level for a structural break model.

	Break - Quarter	mean$(\hat{d})$	Temp. Aggreg.	SB–FDF	Sample Split.	Diff. PP	Diff. KPSS	LWL $\epsilon = 0.02$	LWL $\epsilon = 0.05$
$T = 200$	Small First	0.0800	0.0249	0.2707	0.1331	0.0313	0.0037	0.0064	0.0089
	Small Middle	0.1100	0.0263	0.3112	0.1287	0.0305	0.0024	0.0058	0.0093
	Small Third	0.0800	0.0313	0.2688	0.1274	0.0486	0.0053	0.0047	0.0041
	Medium First	0.2200	0.0307	0.4996	0.1792	0.0274	0.0008	0.0174	0.0266
	Medium Middle	0.2800	0.0191	0.5226	0.1309	0.0241	0.0006	0.0436	0.0606
	Medium Third	0.2200	0.0199	0.5153	0.1800	0.0608	0.0011	0.0167	0.0143
	Large First	0.3600	0.0254	0.5889	0.2464	0.0192	0.0003	0.1282	0.1341
	Large Middle	0.4300	0.0228	0.5624	0.1246	0.0166	0.0007	0.2123	0.2567
	Large Third	0.3700	0.0259	0.5825	0.2606	0.0699	0.0023	0.0396	0.0588
$T = 600$	Small First	0.1324	0.0198	0.4499	0.1689	0.0254	0.0000	0.0099	0.0103
	Small Middle	0.1657	0.0247	0.5732	0.1243	0.0271	0.0014	0.0234	0.0284
	Small Third	0.1307	0.0203	0.4658	0.1594	0.0578	0.0035	0.0071	0.0076
	Medium First	0.3003	0.0302	0.4914	0.3137	0.0130	0.0021	0.1896	0.2260
	Medium Middle	0.3471	0.0262	0.4826	0.1226	0.0104	0.0018	0.3354	0.3904
	Medium Third	0.3003	0.0254	0.5269	0.3801	0.0791	0.0026	0.0795	0.1225
	Large First	0.4319	0.0319	0.4732	0.5326	0.0066	0.0006	0.6413	0.6996
	Large Middle	0.4722	0.0300	0.5655	0.1128	0.0065	0.0007	0.7446	0.7442
	Large Third	0.4276	0.0358	0.6285	0.5899	0.0503	0.0019	0.3317	0.3827
$T = 1000$	Small First	0.1769	0.0228	0.4897	0.2877	0.0187	0.0031	0.0202	0.0462
	Small Middle	0.2068	0.0234	0.6008	0.1303	0.0141	0.0014	0.0703	0.0689
	Small Third	0.1749	0.0206	0.4314	0.2754	0.0434	0.0037	0.0188	0.0200
	Medium First	0.3484	0.0315	0.4492	0.6186	0.0119	0.0000	0.4431	0.6005
	Medium Middle	0.3852	0.0307	0.4505	0.1193	0.0086	0.0018	0.5958	0.7497
	Medium Third	0.3493	0.0343	0.6416	0.6548	0.0521	0.0022	0.2413	0.2815
	Large First	0.4686	0.0374	0.7354	0.7873	0.0087	0.0003	0.9822	0.9637
	Large Middle	0.5010	0.0326	0.7190	0.1642	0.0069	0.0012	0.9663	0.9608
	Large Third	0.4671	0.0397	0.7326	0.8106	0.0616	0.0007	0.8806	0.7451
$T = 2000$	Small First	0.2077	0.0270	0.4303	0.3515	0.0077	0.0008	0.1032	0.0991
	Small Middle	0.2403	0.0256	0.5331	0.1436	0.0121	0.0034	0.2455	0.2670
	Small Third	0.2089	0.0282	0.5094	0.3322	0.0276	0.0033	0.0804	0.0916
	Medium First	0.3772	0.0329	0.5186	0.7108	0.0044	0.0015	0.9399	0.9193
	Medium Middle	0.4080	0.0307	0.4763	0.1396	0.0048	0.0010	0.9151	0.9736
	Medium Third	0.3758	0.0341	0.5112	0.8766	0.0233	0.0046	0.5883	0.6662
	Large First	0.4828	0.0417	0.6607	0.8627	0.0036	0.0012	0.9890	0.9848
	Large Middle	0.5091	0.0398	0.6318	0.2467	0.0078	0.0031	1.0000	1.0000
	Large Third	0.4910	0.0505	0.5669	0.9005	0.0439	0.0000	0.9775	0.9519
$T = 5000$	Small First	0.2143	0.0488	0.4784	0.5947	0.0040	0.0023	0.4739	0.6410
	Small Middle	0.2631	0.0624	0.6849	0.1360	0.0093	0.0108	0.7468	0.7848
	Small Third	0.1968	0.0539	0.3987	0.6212	0.0031	0.0084	0.3364	0.3310
	Medium First	0.3699	0.1017	0.6506	0.8857	0.0682	0.0000	1.0000	0.9874
	Medium Middle	0.4112	0.1842	0.5550	0.1809	0.0827	0.0027	1.0000	0.9779
	Medium Third	0.3723	0.1292	0.7583	0.9089	0.6598	0.0102	1.0000	0.9837
	Large First	0.4866	0.1901	0.6302	0.9466	0.0036	0.0038	1.0000	1.0000
	Large Middle	0.5049	0.2899	0.9001	0.2316	0.0069	0.0016	1.0000	1.0000
	Large Third	0.4830	0.2804	0.6376	0.9111	0.0242	0.0022	1.0000	1.0000

NOTE: See note to Table 3.

TABLE 5: Size-adjusted rejection frequencies at a 5% confidence level for spurious long memory models.

	mean$(\hat{d})$	Temp. Aggreg.	SB–FDF	Sample Split.	Diff. PP	Diff. KPSS	LWL $\epsilon = 0.02$	LWL $\epsilon = 0.05$
				$T = 200$				
RLS-NS	0.1813	0.0213	0.0935	0.1114	0.0057	0.0070	0.0910	0.0998
RLS-S	0.1873	0.0173	0.0889	0.1240	0.0081	0.0105	0.0814	0.0936
STOPBREAK	0.1754	0.0208	0.1366	0.1209	0.0066	0.0201	0.1209	0.1025
MS-IID	0.3613	0.0198	0.2417	0.1174	0.2901	0.0836	0.0885	0.0834
				$T = 600$				
RLS-NS	0.1798	0.0335	0.1360	0.1480	0.0037	0.0491	0.1822	0.2118
RLS-S	0.1369	0.0374	0.1224	0.1378	0.0093	0.0803	0.1801	0.1803
STOPBREAK	0.3403	0.0333	0.3841	0.1241	0.0022	0.2221	0.1415	0.1216
MS-IID	0.4118	0.0389	0.3230	0.0599	0.3413	0.0961	0.1306	0.1101
				$T = 1000$				
RLS-NS	0.1575	0.0401	0.2771	0.1569	0.0090	0.1606	0.2587	0.2366
RLS-S	0.1463	0.0436	0.2509	0.1886	0.0041	0.1398	0.2700	0.2580
STOPBREAK	0.2490	0.0811	0.4993	0.1303	0.0012	0.3617	0.3446	0.3472
MS-IID	0.3930	0.0698	0.3076	0.0898	0.3586	0.1375	0.2102	0.2019
				$T = 2000$				
RLS-NS	0.1212	0.0772	0.5320	0.2002	0.0009	0.2460	0.5313	0.4582
RLS-S	0.1138	0.1456	0.4315	0.2557	0.0036	0.2113	0.5778	0.3709
STOPBREAK	0.3174	0.3005	0.7898	0.2344	0.0015	0.4349	0.7206	0.7700
MS-IID	0.3404	0.1743	0.1987	0.1056	0.2803	0.1986	0.5821	0.5631
				$T = 5000$				
RLS-NS	0.2465	0.3528	0.7317	0.3307	0.0006	0.6924	0.8536	0.8427
RLS-S	0.1952	0.4719	0.5981	0.4305	0.0028	0.5890	0.9202	0.8650
STOPBREAK	0.3324	0.8312	0.7903	0.3110	0.0029	0.6854	0.9555	0.9793
MS-IID	0.2356	0.3557	0.0566	0.1066	0.1972	0.4309	0.9125	0.9309

NOTE: See note to Table 3. In addition, RLS-NS denotes the Nonstationary random level shift model, RLS-S is the stationary random level shift model, STOPBREAK is the model of Engle and Smith (1999) and MS-IID is the Markov-Switching model.

TABLE 6: Rejection frequencies at a 5% confidence level for a structural break model.

	Break - Quarter	mean$(\hat{d})$	Temp. Aggreg.	SB–FDF	Sample Split.	Diff. PP	Diff. KPSS	LWL $\epsilon = 0.02$	LWL $\epsilon = 0.05$
$T = 200$	Small First	0.0800	0.0206	0.2390	0.0794	0.0169	0.0183	0.0390	0.0355
	Small Middle	0.1100	0.0220	0.2795	0.0750	0.0177	0.0196	0.0396	0.0351
	Small Third	0.0800	0.0270	0.2371	0.0737	0.0004	0.0167	0.0407	0.0403
	Medium First	0.2200	0.0264	0.4679	0.1255	0.0208	0.0212	0.0280	0.0178
	Medium Middle	0.2800	0.0148	0.4909	0.0772	0.0241	0.0214	0.0018	0.0162
	Medium Third	0.2200	0.0156	0.4836	0.1263	0.0126	0.0209	0.0287	0.0301
	Large First	0.3600	0.0211	0.5572	0.1927	0.0290	0.0217	0.0828	0.0897
	Large Middle	0.4300	0.0185	0.5307	0.0709	0.0316	0.0213	0.1669	0.2123
	Large Third	0.3700	0.0216	0.5508	0.2069	0.0217	0.0197	0.0058	0.0144
$T = 600$	Small First	0.1324	0.0145	0.4162	0.1131	0.0258	0.0234	0.0384	0.0370
	Small Middle	0.1657	0.0194	0.5395	0.0685	0.0241	0.0220	0.0249	0.0189
	Small Third	0.1307	0.0150	0.4321	0.1036	0.0066	0.0199	0.0412	0.0397
	Medium First	0.3003	0.0249	0.4577	0.2579	0.0382	0.0213	0.1413	0.1787
	Medium Middle	0.3471	0.0209	0.4489	0.0668	0.0408	0.0216	0.2871	0.3431
	Medium Third	0.3003	0.0201	0.4932	0.3243	0.0279	0.0208	0.0312	0.0752
	Large First	0.4319	0.0266	0.4395	0.4768	0.0446	0.0228	0.5930	0.6523
	Large Middle	0.4722	0.0247	0.5318	0.0570	0.0447	0.0227	0.6963	0.6969
	Large Third	0.4276	0.0305	0.5948	0.5341	0.0009	0.0215	0.2834	0.3354
$T = 1000$	Small First	0.1769	0.0173	0.4573	0.2299	0.0306	0.0194	0.0263	0.0007
	Small Middle	0.2068	0.0179	0.5684	0.0725	0.0352	0.0211	0.0238	0.0235
	Small Third	0.1749	0.0151	0.3990	0.2176	0.0059	0.0188	0.0277	0.0255
	Medium First	0.3484	0.0260	0.4168	0.5608	0.0374	0.0225	0.3966	0.5551
	Medium Middle	0.3852	0.0252	0.4181	0.0615	0.0407	0.0207	0.5493	0.7043
	Medium Third	0.3493	0.0288	0.6092	0.5970	0.0028	0.0203	0.1948	0.2361
	Large First	0.4686	0.0319	0.7030	0.7295	0.0406	0.0222	0.9357	0.9183
	Large Middle	0.5010	0.0271	0.6866	0.1064	0.0424	0.0213	0.9198	0.9154
	Large Third	0.4671	0.0342	0.7002	0.7528	0.0123	0.0218	0.8341	0.6997
$T = 2000$	Small First	0.2077	0.0213	0.3972	0.2916	0.0427	0.0222	0.0557	0.0526
	Small Middle	0.2403	0.0199	0.5000	0.0837	0.0383	0.0196	0.1980	0.2205
	Small Third	0.2089	0.0225	0.4763	0.2723	0.0228	0.0197	0.0329	0.0451
	Medium First	0.3772	0.0272	0.4855	0.6509	0.0460	0.0215	0.8924	0.8728
	Medium Middle	0.4080	0.0250	0.4432	0.0797	0.0456	0.0220	0.8676	0.9271
	Medium Third	0.3758	0.0284	0.4781	0.8167	0.0271	0.0184	0.5408	0.6197
	Large First	0.4828	0.0360	0.6276	0.8028	0.0468	0.0218	0.9415	0.9383
	Large Middle	0.5091	0.0341	0.5987	0.1868	0.0426	0.0199	0.9525	0.9535
	Large Third	0.4910	0.0448	0.5338	0.8406	0.0065	0.0230	0.9300	0.9054
$T = 5000$	Small First	0.2143	0.0441	0.4438	0.5328	0.0486	0.0217	0.4243	0.5925
	Small Middle	0.2631	0.0577	0.6503	0.0741	0.0433	0.0132	0.6972	0.7363
	Small Third	0.1968	0.0492	0.3641	0.5593	0.0495	0.0156	0.2868	0.2825
	Medium First	0.3699	0.0970	0.6160	0.8238	0.0156	0.0240	0.9504	0.9389
	Medium Middle	0.4112	0.1795	0.5204	0.1190	0.0301	0.0213	0.9505	0.9294
	Medium Third	0.3723	0.1245	0.7237	0.8470	0.6072	0.0138	0.9614	0.9352
	Large First	0.4866	0.1854	0.5956	0.8847	0.0490	0.0202	0.9578	0.9515
	Large Middle	0.5049	0.2852	0.8655	0.1697	0.0457	0.0224	0.9531	0.9531
	Large Third	0.4830	0.2757	0.6030	0.8492	0.0284	0.0218	0.9554	0.9529

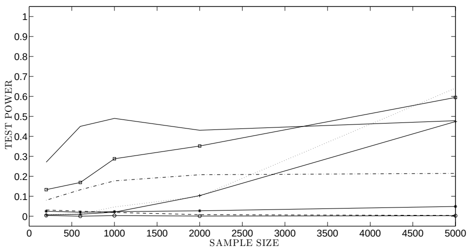
NOTE: See note to Table 3.

TABLE 7: Rejection frequencies at a 5% confidence level for spurious long memory models.

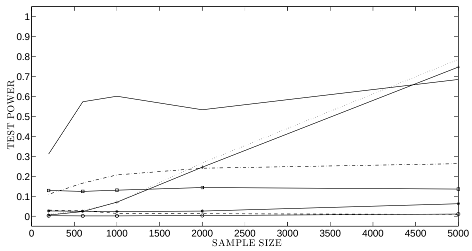
	mean$(\hat{d})$	Temp. Aggreg.	SB–FDF	Sample Split.	Diff. PP	Diff. KPSS	LWL $\epsilon = 0.02$	LWL $\epsilon = 0.05$
				$T = 200$				
RLS-NS	0.1813	0.0170	0.0618	0.0577	0.0425	0.0150	0.0456	0.0554
RLS-S	0.1873	0.0130	0.0572	0.0703	0.0401	0.0115	0.0360	0.0492
STOPBREAK	0.1754	0.0165	0.1049	0.0672	0.0416	0.0019	0.0755	0.0581
MS-IID	0.3613	0.0793	0.2584	0.0484	0.0474	0.0135	0.0650	0.0548
				$T = 600$				
RLS-NS	0.1798	0.0282	0.1023	0.0922	0.0475	0.0257	0.1339	0.1645
RLS-S	0.1369	0.0321	0.0887	0.0820	0.0419	0.0569	0.1318	0.1330
STOPBREAK	0.3403	0.0280	0.3504	0.0683	0.0490	0.1987	0.0932	0.0743
MS-IID	0.4118	0.0908	0.3076	0.0431	0.1986	0.0218	0.1293	0.1187
				$T = 1000$				
RLS-NS	0.1575	0.0346	0.2447	0.0991	0.0403	0.1381	0.2122	0.1912
RLS-S	0.1463	0.0381	0.2185	0.1308	0.0452	0.1173	0.2235	0.2126
STOPBREAK	0.2490	0.0756	0.4669	0.0725	0.0481	0.3392	0.2981	0.3018
MS-IID	0.3930	0.1320	0.3262	0.0459	0.3061	0.0194	0.2854	0.2674
				$T = 2000$				
RLS-NS	0.1212	0.0715	0.4989	0.1403	0.0495	0.2230	0.4838	0.4117
RLS-S	0.1138	0.1399	0.3984	0.1958	0.0468	0.1883	0.5303	0.3244
STOPBREAK	0.3174	0.2948	0.7567	0.1745	0.0489	0.4119	0.6731	0.7235
MS-IID	0.3404	0.1929	0.2472	0.0607	0.3614	0.0207	0.5945	0.5481
				$T = 5000$				
RLS-NS	0.2465	0.3481	0.6971	0.2688	0.0520	0.6684	0.8040	0.7942
RLS-S	0.1952	0.4672	0.5635	0.3686	0.0498	0.5650	0.8706	0.8165
STOPBREAK	0.3324	0.8265	0.7557	0.2491	0.0497	0.6614	0.9059	0.9308
MS-IID	0.2356	0.4262	0.1626	0.0624	0.5576	0.0190	0.8541	0.8687

NOTE: See note to Table 5.

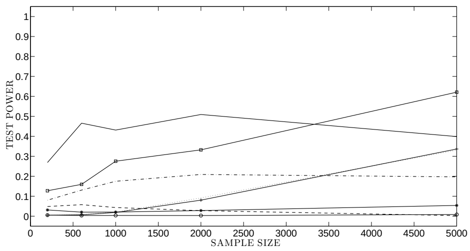
FIGURE 1: Size-adjusted power of the tests in the structural break model: Small break, bandwidth $m = \left[T^{0.5}\right]$.



(a) First quarter



(b) Middle



(c) Third quarter

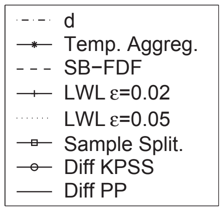
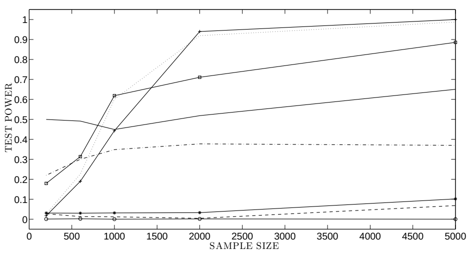
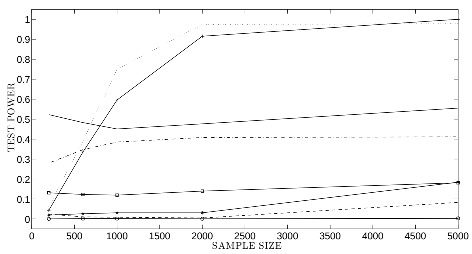


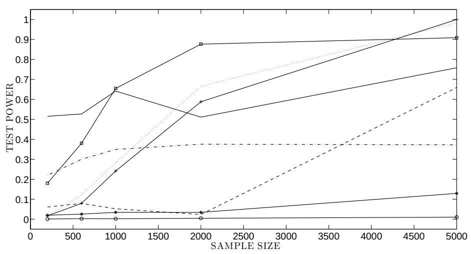
FIGURE 2: Size-adjusted power of the tests in the structural break model: Medium break, bandwidth $m = [T^{0.5}]$.



(a) First quarter



(b) Middle



(c) Third quarter

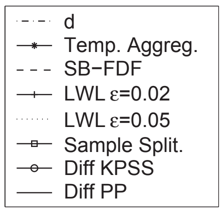
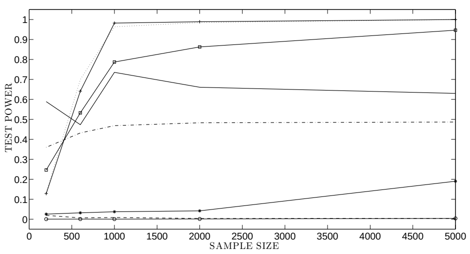
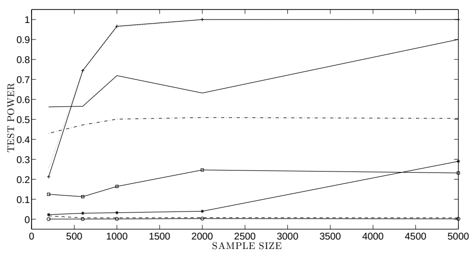


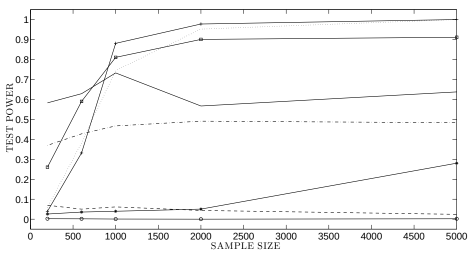
FIGURE 3: Size-adjusted power of the tests in the structural break model: Large break, bandwidth $m = \left[T^{0.5}\right]$.

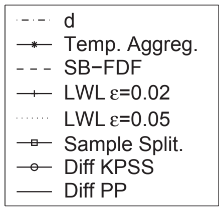


(a) First quarter



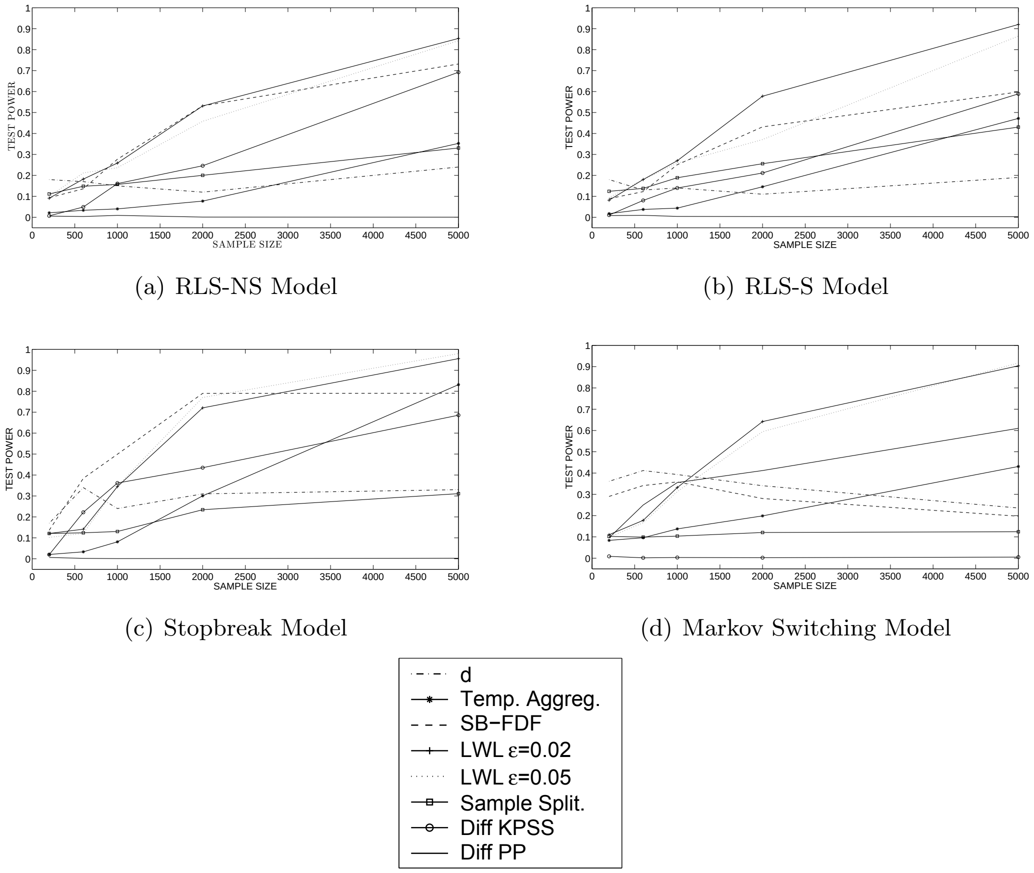
(b) Middle





(c) Third quarter

FIGURE 4: Size-adjusted power of the tests for spurious long memory models, bandwidth $m = \left[T^{0.5}\right]$.



(a) RLS-NS Model

(b) RLS-S Model

(c) Stopbreak Model

(d) Markov Switching Model