Can Markov switching model generate long memory?

Changryong Baek [a,*], Natércia Fortuna [b,1], Vladas Pipiras [c,2]

[a] *Sungkyunkwan University, South Korea*
[b] *CEF.UP - Centre for Economics and Finance at University of Porto, Universidade do Porto, Portugal*
[c] *University of North Carolina, United States*

HIGHLIGHTS

- The relationship between Markov switching (MS) model and long memory is reexamined.
- Common spectral estimators of the long memory parameter are found to be extremely biased for the MS model.
- This is explained by analyzing the expected periodogram and the variance of the partial sum process of the MS model.

ARTICLE INFO

ABSTRACT

In an influential work by Diebold and Inoue (2001), the Markov switching model was shown to exhibit long memory, in terms of the behavior of the second moments of partial sums. The relationship between the Markov switching model and long memory is reexamined here. Common estimators of the long memory parameter are found to be extremely biased when applied to the data generated by the Markov switching model. An explanation for these findings is provided.

© 2014 Elsevier B.V. All rights reserved.

1. Introduction

In a seminal work, Diebold and Inoue (2001) related several forms of changes in regime to long memory. One of several fundamental models studied in their work was the Markov switching (MS) model, defined as

$$X_k^T = \mu_{s_k^T} + \epsilon_k, \quad k = 1, \ldots, T, \tag{1.1}$$

* Correspondence to: Department of Statistics, Sungkyunkwan University, 25-2, Sungkyunkwan-ro, Jongno-gu, Seoul, 110-745, South Korea. Tel.: +82 2 760 0602; fax: +82 502 302 0522.

E-mail addresses: crbaek@skku.edu, crbaek@gmail.com (C. Baek), nfortuna@fep.up.pt (N. Fortuna), pipiras@email.unc.edu (V. Pipiras).

[1] Address: Faculdade de Economia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-464 Porto, Portugal.
[2] Address: Department of Statistics and Operations Research, UNC at Chapel Hill, CB#3260, Hanes Hall, Chapel Hill, NC 27599, USA.

where $\{\epsilon_k\}$ is a sequence of i.i.d. random variables with mean 0 and variance σ_ϵ^2, $\{s_k^T\}_{k=1,\ldots,T}$ is a stationary Markov chain, taking the value 0 or 1, with the transition probabilities $p_{ij} = p_{ij}(T)$, $i, j = 0, 1$, depending on T, and $\mu_0 \neq \mu_1$ are two different mean levels.

Supposing

$$p_{jj} = p_{jj}(T) = 1 - \frac{c_j}{T^{\delta_j}}, \quad j = 0, 1 \tag{1.2}$$

with $0 < c_0, c_1 < 1$ and $0 < \delta_0, \delta_1 < 1$, Diebold and Inoue (2001, Proposition 3), showed that

$$\text{Var}(S_T) \sim CT^{2d+1}, \quad \text{as } T \to \infty, \tag{1.3}$$

where $S_T = X_1^T + \cdots + X_T^T$ is the partial sum of the series X_k^T, and $d = (\min\{\delta_0, \delta_1\} - |\delta_0 - \delta_1|)/2 \in (0, 1/2)$.

The behavior (1.3) of the variances of the partial sums is consistent with that of long memory time series, with the long memory parameter $d \in (0, 1/2)$ (e.g. Robinson, 2003; Beran et al., 2013). The MS model thus seemingly suggests a way to generate (induce)

long memory, providing another physical mechanism for long memory akin to ON/OFF models (Willinger et al., 1997; Mikosch et al., 2002; Smith, 2005) or aggregation of short memory models (Granger, 1980).

In this note, we reexamine the relationship between the MS model and long memory. We find through numerical simulations that the spectral (GPH or local Whittle) estimation of the long memory parameter d in the MS model is extremely biased and very sensitive to the tuning parameter (the number of low frequencies used in estimation). This occurs because the population quantities (the log of the expected periodogram) deviate very significantly from what would be expected under long memory (see Section 2).

We shed further light on the relationship between the MS model and long memory by examining the variance of the partial sum process $S_T(t) = \sum_{k=1}^{[Tt]}(X_k^T - EX_k^T)$, $t \in [0, 1]$. As shown below (see Section 3), the behavior of this variance is more in line with short memory (associated with $d = 0$) and not long memory ($d \in (0, 1/2)$ as above). Since this behavior often underlies the estimation of the long (short) memory parameter, the MS model should thus be viewed as a short memory model which happens to have some long memory features as in (1.3). This also sheds light on the estimation results in the spectral domain.

Our findings show that the MS model should not be viewed as another physical mechanism to generate true long memory (see also Section 3 below). The MS model may nevertheless suggest long memory through common spectral estimators such as GPH or local Whittle. Diebold and Inoue (2001) showed this by focusing on testing for short memory (which can be rejected for the MS model). We differ from Diebold and Inoue (2001) by focusing on point estimates of the long memory parameter. This not only confirms the findings of Diebold and Inoue (2001) but also identifies the issue of a serious bias in estimating the expected long memory parameter.

The rest of the note is organized as follows. Estimation results are presented in Section 2 and further discussion can be found in Section 3. Section 4 contains some concluding remarks.

2. Estimation of long memory parameter

In this section, we consider the estimation of the long memory parameter in the MS model. We first consider the popular GPH and local Whittle (LW, in short) methods. Both methods are in the spectral domain, and use the periodogram $I_X(\omega_j)$ of the series $\{X_k^T\}$ at the Fourier frequencies $\omega_j = 2\pi j/T$, $j = 1, \ldots, T$. The GPH estimator is obtained from the regression of $\log I_X(\omega_j)$ on $\log \omega_j$ and can be expressed as

$$\widehat{d}_{gph} = \sum_{j=1}^{m} a_j \log I_X(\omega_j),$$

where m is the number of low frequencies used in estimation and the regression weights $a_j = (z_j - \bar{z})/\sum_{j=1}^{m}(z_j - \bar{z})^2$ with $z_j = -2\log\omega_j$, $\bar{z} = \sum_{j=1}^{m} z_j/m$ (Geweke and Porter-Hudak, 1983). The LW estimator, based on the Whittle approximation of the Gaussian log-likelihood, is defined as (Robinson, 1995)

$$\widehat{d}_{lw} = \underset{d \in [\Theta_1, \Theta_2]}{\operatorname{argmin}} \left\{ \log\left(\frac{1}{m}\sum_{j=1}^{m}\omega_j^{2d} I_X(\omega_j)\right) - \frac{2d}{m}\sum_{j=1}^{m}\log\omega_j \right\}, \quad (2.1)$$

where $-0.5 < \Theta_1 < 0 < \Theta_2 < 1$ are fixed.

Fig. 1 depicts GPH and LW estimation for the MS model. We took $\mu_0 = 0$, $\mu_1 = 1$, $c_0 = c_1 = 0.9$, normal errors with $\sigma_\epsilon^2 = 1$, $T = 10^5$ and two choices of $\delta_0 = \delta_1 = 0.8$ and $\delta_0 = \delta_1 = 0.2$ corresponding to $d = 0.4$ and $d = 0.1$, respectively. The top plots are the estimates $\widehat{d} = \widehat{d}(m)$ as functions of m when $d = 0.4$, for 100 realizations. The bottom plots correspond to $d = 0.1$. It is clearly

seen from Fig. 1 that both estimators are extremely biased for a wide range of low frequencies used in estimation even for such a large sample size. For larger $d = 0.4$, the estimators suggest long memory for a wide range of lower frequencies but the bias is close to zero only for a tiny subset of this range.

The difficulties with estimation can be explained further through the following argument. By using Proposition 3 in Diebold and Inoue (2001), which provides the autocovariance function of the MS model $\{X_k^T\}$, it can be shown that

$$\mathbb{E}I_X(\omega_j) = \frac{1}{2\pi}\left\{\sigma_\epsilon^2 + \frac{(1-p_{00})(1-p_{11})}{(2-p_{00}-p_{11})^2}(\mu_0 - \mu_1)^2\right.$$
$$\times \left\{1 + 2\sum_{k=1}^{T-1}\left(1 - \frac{|k|}{T}\right)\lambda^k \cos(k\omega_j)\right\}\right\}$$
$$= \frac{1}{2\pi}\left\{\sigma_\epsilon^2 + \frac{(1-p_{00})(1-p_{11})}{(2-p_{00}-p_{11})^2}(\mu_0 - \mu_1)^2\right.$$
$$\times \left\{1 + 2\Re\left(\frac{\lambda e^{i\omega_j}}{1 - \lambda e^{i\omega_j}}\right.\right.$$
$$\left.\left.\left. + \frac{1}{T}\frac{(\lambda e^{i\omega_j})^{T+1} - \lambda e^{i\omega_j}}{(1 - \lambda e^{i\omega_j})^2}\right)\right\}\right\}, \quad (2.2)$$

where $\Re$ denotes the real part. Fig. 2 shows the log–log plot of the expected periodogram (2.2) in plus marker, an average of 100 periodograms from simulated data in dotted line, and the slope with the long memory parameter d in dashed–dotted line, against the Fourier frequencies. Note the good agreement between the expected and averaged periodograms. But also note that they deviate very significantly from the line of the expected slope determined by d.

Diebold and Inoue (2001) have also reported some numerical simulations for the MS model (see also Yu, 2009). In figures 8 and 9, they consider the GPH estimates for the choice of $m = T^{0.5}$, finding them in the interval $(0, 0.5)$. Although these findings are consistent with the ones reported here (Fig. 1), we go a step further by focusing on the bias of the GPH estimator.

3. Behavior of variance of partial sum process

A disagreement of the MS model with the expected long memory behavior also manifests itself in the following more basic and fundamental way. If $S_T(t) = \sum_{k=1}^{[Tt]}(X_k^T - EX_k^T)$, $t \in [0, 1]$, is the partial sum process of the MS model, then Proposition 3 in Diebold and Inoue (2001) can be used to show that

$$\operatorname{Var}(S_T(t)) = \frac{(1-p_{00})(1-p_{11})}{(2-p_{00}-p_{11})^2}(\mu_0 - \mu_1)^2[Tt]$$
$$\times \left\{\frac{2\lambda}{1-\lambda} + 1 + \frac{2(\lambda^{[Tt]+1} - \lambda)}{[Tt](1-\lambda)^2}\right\} + \sigma_\epsilon^2[Tt]. \quad (3.1)$$

This further leads to

$$v_T(t) := \operatorname{Var}\left(\frac{S_T(t)}{T^{d+1/2}}\right) \sim Ct, \quad (3.2)$$

as $T \to \infty$ (see Baek et al., 2014 for details).

In the case of a true long memory time series, the right-hand side of (3.2) is Ct^{2d+1}. The behavior (3.2), on the other hand, is associated with short memory and $d = 0$. (As shown in Baek et al., 2014, $S_T(t)/T^{d+1/2}$ also converges in the sense of finite-dimensional distributions to Brownian motion.) Fig. 3 presents the log–log plot of $v_T(t)$ as a function of t, together with the expected slope equal to 1 ($d = 0$). Note the good agreement between $\log v_T(t)$ and the expected slope.

In fact, the relation (3.2) with t^{2d+1} on its right-hand side gives rise to the so-called aggregated variance estimator of the long
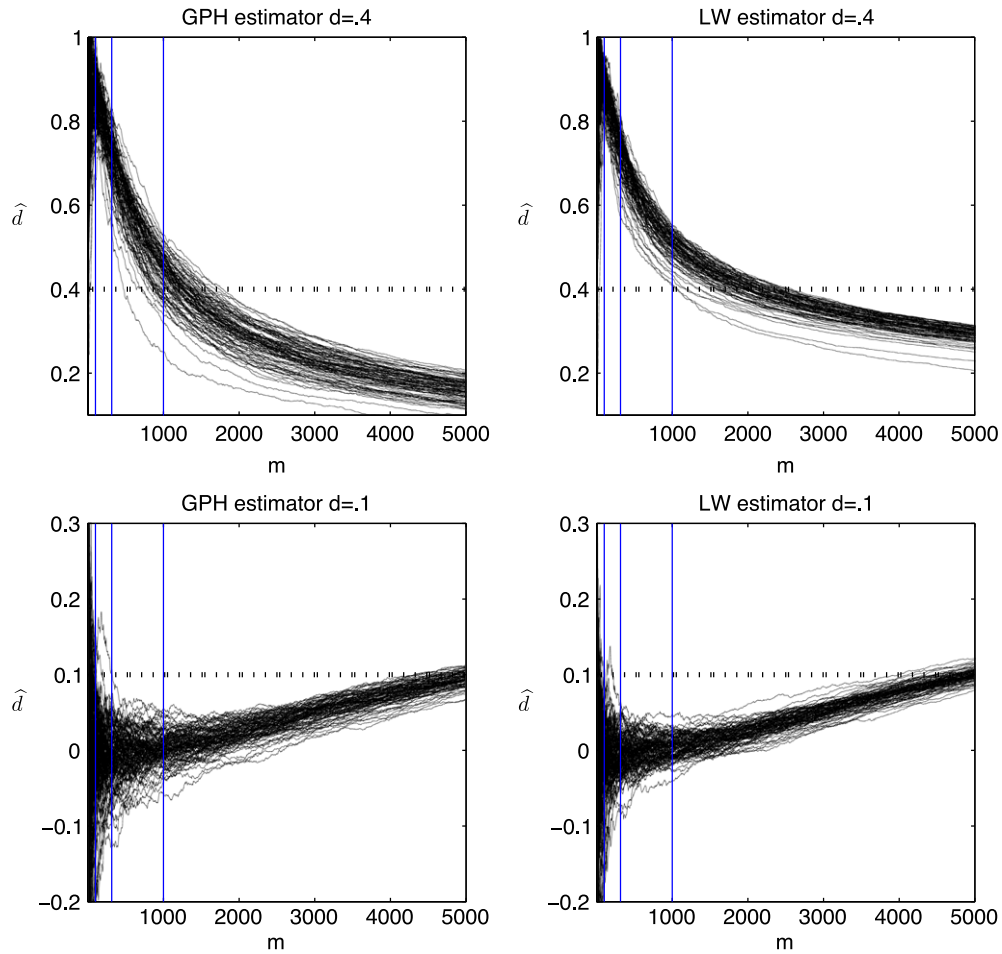
Fig. 1. GPH and LW estimation for the MS model with $d = 0.4$ and 0.1. The dashed horizontal line indicates the value of $d = 0.4$ or 0.1. Three vertical lines correspond to $m = T^{0.4}$, $T^{0.5}$ and $T^{0.6}$, as rule-of-thumb choices for m.
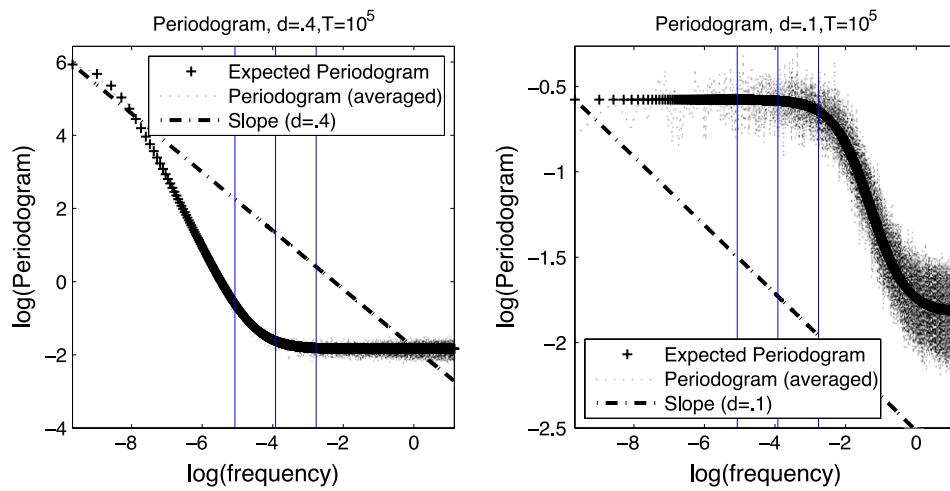

Fig. 2. Periodogram against Fourier frequencies in the log–log plots. The slope of the dashed–dotted line is determined by the true long memory parameter.

memory parameter d, which according to (3.2) is zero for the MS model, corresponding to short memory. From the estimation perspective, the MS model thus behaves more like a short memory model which happens to have some long memory features as in (1.3). This claim also sheds light on the estimation result in the spectral domain: the slope of the expected periodogram in Fig. 2 tends to have the zero slope at lowest frequencies (for a larger d, a much larger sample size is needed to observe the zero slope at a larger range of the lowest frequencies).

4. Conclusion

In this note, we reexamined the connection between the Markov switching model and long memory. The theoretical results of Diebold and Inoue (2001) relate the model to long memory, but a serious bias makes the expected long memory parameter difficult to observe in practice. By examining the variance of the partial sum process, we also argued that the MS model should be viewed rather as a short memory model which happens to have some long
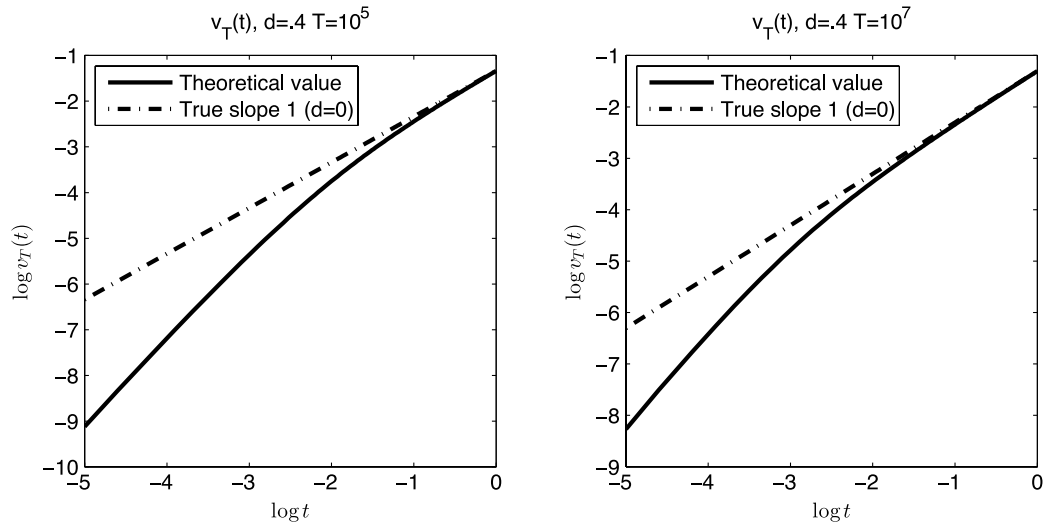
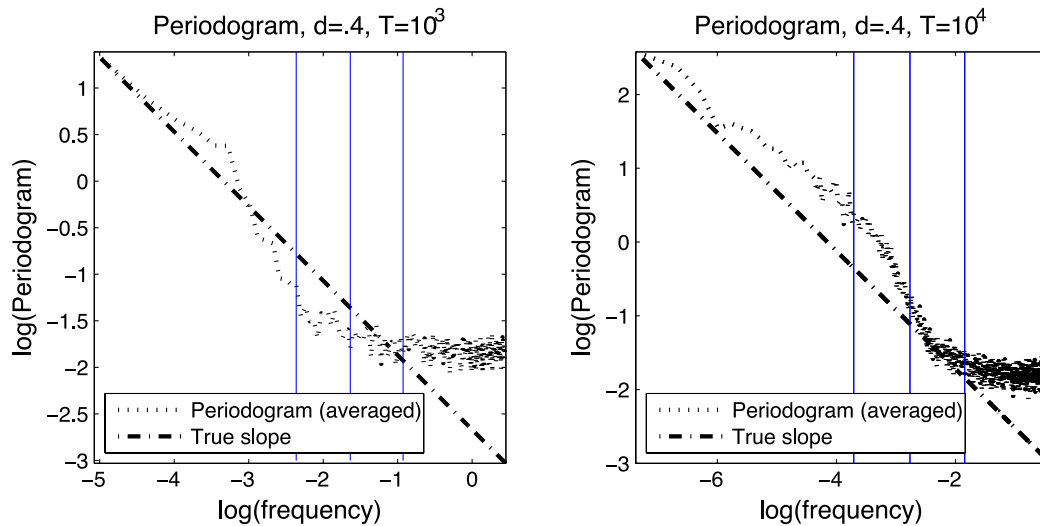**Fig. 3.** $\log v_T(t)$ against $\log t$ for the MS model with $d = 0.4$.

Fig. 4. Log–log plots for the ON/OFF model with $d = 0.4$ and sample sizes $T = 10^3$ and $T = 10^4$.

memory features. In particular, the MS model may suggest long memory through common estimators in the spectral domain.

Related to this last point, we also mention several recent works aiming at distinguishing changes-in-mean and related models from long memory. For example, Iacone (2010), McCloskey and Perron (2013) employ common estimators of long memory in the spectral domain by selecting carefully the range of frequencies to be used in estimation. Baek and Pipiras (2012, 2013) propose a method based on removing changes in mean and local Whittle estimation.

Our findings for the MS model also stand in sharp contrast to other physical mechanisms to generate long memory. For example, the ON/OFF model (superimposed with i.i.d. noise) can be defined by (1.1) where $s_k^T = s_k$ does not depend on T and still takes values 0 or 1 but now

$$s_k = \begin{cases} 1, & \text{if } S_i \le k < S_i + U_i^{\text{ON}}, \\ 0, & \text{if } S_i + U_i^{\text{ON}} \le k < S_{i+1}, \end{cases}$$

with $S_i = S_0 + \sum_{j=1}^{i}(U_j^{\text{ON}} + U_j^{\text{OFF}})$ and independent collections $\{U_j^{\text{ON}}\}$ and $\{U_j^{\text{OFF}}\}$ of i.i.d. positive random variables. (S_0 is a special random variable to ensure that s_k is stationary.) Such an ON/OFF model is long memory as long as the distribution $F = F_{\text{ON}}$ of $\{U_j^{\text{ON}}\}$

or $F = F_{\text{OFF}}$ of $\{U_j^{\text{OFF}}\}$ is heavy-tailed, that is, $1 - F(u) \sim cu^{-\alpha}$, as $u \to \infty$, where $\alpha \in (1, 2)$. The long memory parameter d is related to α through $d = (2 - \alpha)/2$.

Fig. 4 presents the log–log plots of the averaged periodogram against Fourier frequencies for the ON/OFF model where ON times are chosen to be Pareto with $\alpha = 1.2$ (corresponding to $d = 0.4$), $c = 50^{1.2}$, and OFF times follow the exponential distribution with mean 150. The noise variance is $\sigma_\epsilon^2 = 1$ as in the MS model simulations, and the results are based on 100 replications. As seen from the plots, the log of the periodogram exhibits a linear behavior with the slope close to the true one for a number of low frequencies considered.

Acknowledgments

References

Baek, C., Fortuna, N., Pipiras, V., 2014. Convergence of the long memory Markov switching model to Brownian motion, Technical Note. Available at http://www.stat.unc.edu/faculty/pipiras.

Baek, C., Pipiras, V., 2012. Statistical tests for a single change in mean against long-range dependence. J. Time Ser. Anal. 33 (1), 131–151.

Baek, C., Pipiras, V., 2013. On distinguishing multiple changes in mean and long-range dependence using local Whittle estimation, Preprint. Available at http://www.stat.unc.edu/faculty/pipiras.

Beran, J., Feng, Y., Ghosh, S., Kulik, R., 2013. Long-Memory Processes Probabilistic Properties and Statistical Methods. Springer, Heidelberg.

Diebold, F.X., Inoue, A., 2001. Long memory and regime switching. J. Econometrics 105 (1), 131–159.

Geweke, J., Porter-Hudak, S., 1983. The estimation and application of long memory time series models. J. Time Ser. Anal. 4 (4), 221–238.

Granger, C.W.J., 1980. Long memory relationships and the aggregation of dynamic models. J. Econometrics 14 (2), 227–238.

Iacone, F., 2010. Local Whittle estimation of the memory parameter in presence of deterministic components. J. Time Ser. Anal. 31 (1), 37–49.

McCloskey, A., Perron, P., 2013. Memory parameter estimation in the presence of level shifts and deterministic trends. Econometric Theory 29 (06), 1196–1237.

Mikosch, T., Resnick, S., Rootzén, H., Stegeman, A., 2002. Is network traffic approximated by stable Lévy motion or fractional Brownian motion? Ann. Appl. Probab. 12 (1), 23–68.

Robinson, P.M., 1995. Gaussian semiparametric estimation of long range dependence. Ann. Statist. 23 (5), 1630–1661.

Robinson, P.M. (Ed.), 2003. Time Series With Long Memory. In: Advanced Texts in Econometrics, Oxford University Press, Oxford.

Smith, A., 2005. Level shifts and the illusion of long memory in economic time series. J. Bus. Econom. Statist. 23 (3), 321–335.

Willinger, W., Taqqu, M.S., Sherman, R., Wilson, D.V., 1997. Self-similarity through high-variability: statistical analysis of ethernet LAN traffic at the source level. IEEE/ACM Trans. Netw. 5 (1), 71–86.

Yu, W.-C., 2009. Markov switching and long memory: a Monte Carlo analysis. Appl. Econ. Lett. 16 (12), 1205–1210.