Supplementary material on "Signal Extraction in Long Memory Stochastic Volatility"

Josu Arteche *
Dept. of Econometrics and Statistics
University of the Basque Country UPV/EHU
Bilbao 48015
Spain
email: josu.arteche@ehu.es

Revised: 5th August 2014

This supplementary material contains a detailed Monte Carlo analysis comparing our proposal for signal extraction with two natural competitors adapted to the semiparametric character of the problem at hand: a Wiener-Kolmogorov filter in the time domain as proposed by Harvey (1998) and smoothing via the Kalman filter in a truncated AR process. The applicability of our proposal is finally illustrated in an empirical analysis of a daily series of returns from the Dow Jones Industrial index.

1 Finite sample performance

We compare the performance of our proposal with two extensions of existing techniques for signal extraction in SV models: the Kalman filter, which is the most widely used tool for estimating the volatility in parametric short memory SV models, and the proposal by Harvey (1998) for parametric LMSV based on a Wiener-Kolmogorov filter in the time domain.

1.1 Kalman filter in LMSV

SV models can be naturally expressed in state space form and the Kalman filter can be used to construct the likelihood function and to extract the volatility component. This approach gives reliable results with short memory volatility components, but a strong persistent x_t poses difficulties which may render the Kalman filter quite unreliable and unmanageable

*Research supported by the Spanish Ministry of Science and Innovation and ERDF grants ECO2010-15332 and ECO2013-40935-P, UPV/EHU Econometrics Research Group, Basque Government grant IT-642-13 and UPV/EHU UFI 11/03 Sustainable Economics & Welfare.

due to the huge dimension of the corresponding state space model. In fact, Chan and Palma (1998) show that the state space representation of a long memory process cannot be finite dimensional, though they do not deal with SV models. The exact likelihood function can however be computed in a finite number of steps, but the computation may be rather cumbersome, with a number of computations of order n^3 for n denoting the sample size, which makes it quite unmanageable for the sample sizes usually found in financial series. Moreover its application requires a full parameterization of x_t including its short memory part, and this is a restriction that we wish to avoid. A partial solution to this problem is to work with a truncated MA or AR expression as suggested by Chan and Palma (1998). This reduces the number of operations required for a single evaluation of the likelihood function to the order of n.

Truncating the MA expansion gives rise to the inconvenience of a very slow decay of the MA coefficients. In a long memory set up, the lag-j coefficient in the MA expansion is proportional to j^{d-1}, which implies that the truncating point needs to be quite large if serious problems of misspecification are to be avoided. Chan and Palma (1998) suggest instead truncating the first differences of the series such that the lag-j coefficient of the MA expansion is now of order j^{d-2} and the truncation can be executed with fewer components. This strategy performs well for estimating a parametric Fractional ARIMA process, as suggested by Chan and Palma (1998), but its application to signal extraction in LMSV models as defined in the main text poses certain problems. First, in contrast with parametric models where the number of parameters to be estimated remains fixed independently of the truncation point, in our local or semiparametric context the number of parameters increases with the truncation point. Second, taking first differences implies an undesirable transformation of the signal which has to be reversed to get estimates of the original signal. This is not exact in finite samples and depends on the initial values selected. Third, the added noise in the measurement equation of the differenced model is no longer white noise but a noninvertible $MA(1)$.

For signal extraction in local LMSV models such as those discussed here we have found it more suitable to truncate an AR expansion of the original series and to estimate the volatility component by smoothing using the Kalman filter. The advantages of this approach are that it allows a lower truncation (which implies fewer parameters to be estimated) because the AR coefficients decrease faster to zero (proportional to j^{-d-1}), it needs no prior transformations of the data and it does not affect the white noise character of the added noise. However, the approach suffers from problems caused by the misspecification of the long memory signal, whereas our proposal incorporates this characteristic into the definition of the weights of the filter.

1.2 Wiener-Kolmogorov filter in the time domain (Harvey)

Harvey (1998) proposes estimating a stationary x_t by applying a linear Wiener-Kolmogorov filter that minimizes the mean square error (MSE). Under the assumptions in the paper it takes the form

$$\tilde{x} = (I - \sigma_u^2 \Sigma_y^{-1})(y - \mu) \tag{1}$$

where Σ_y is the variance covariance matrix of y and σ_u^2 is the variance of the added noise. The empirical implementation of this signal extraction strategy suffers from some serious drawbacks. First it requires inversion of Σ_y, which can be very computationally demanding if the sample size is large. Moreover, due to the persistent autocorrelation Σ_y may be close to being singular and its inverse may be rather unstable. Secondly, unknowns have to be estimated and the quality of the estimates significantly affects the signal extraction, as evidenced by the results in this Monte Carlo analysis. Thirdly, it is only valid for stationary series. In a non stationary context Harvey (1998) suggests prior differencing such that the added noise loses its white noise characteristic and the original signal is estimated by integrating the estimated differenced signal as explained and implemented in the Monte Carlo analysis below.

1.3 Monte Carlo analysis

The finite sample performance of the signal extraction methods is analyzed in 1000 replications of series generated as

$$y_t = x_t + u_t \quad t = 1, 2, ..., n,$$

for $x_t = \kappa x_t^*$, $(1 - L)^{d_0} x_t^* = w_t$ and six different specifications are considered for signal and noise:

Model 1 : $d_0 = 0.4$, $w_t = w_t^*$ and $u_t = \log \epsilon_t^2$ with

$$\begin{pmatrix} \epsilon_t \\ w_{t-1}^* \end{pmatrix} \sim NID \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix} \right]$$

Model 2 : Same as Model 1 but with $(1 - 0.8L)w_t = w_t^*$.

Model 3 : Same as Model 1 but with $(1 - 0.2L + 0.8L^2)w_t = w_t^*$.

Model 4 : Same as Model 1 but with $d_0 = 0.8$.

Model 5 : $d_0 = 0.4$ and $(w_t, u_t)' = H_t^{1/2}\eta_t$ for $\eta_t \sim N(0, I_2)$, I_2 is the identity matrix of dimension 2, $H_t = diag(a_1 h_{1t}, a_2 h_{2t})$, $h_{it} = \alpha_0 + \alpha_1 w_{t-1}^2 + \alpha_2 u_{t-1}^2$ for $i = 1, 2$, $\alpha = (\alpha_0, \alpha_1, \alpha_2) = (0.0001, 0.25, 0.04)$ and a_1, a_2 are constants chosen to maintain the unconditional variances of signal and noise as in Model 1.

Model 6 : $d_0 = 0.4$, $u_t = \log \epsilon_t^2$ and $w_t = g(\epsilon_{t-1})/\sqrt{var(g(\epsilon_{t-1}))}$ with $g(\epsilon_t) = 0.3(|\epsilon_t| - \sqrt{2/\pi}) - 0.2\epsilon_t$ for $\epsilon_t \sim \mathcal{NID}(0, 1)$ such that $var(g(\epsilon_{t-1})) = 0.0727$.

Model 1 corresponds to a stationary signal with spectral power concentration around the origin. Two different values of ρ are considered, $\rho = 0, -0.8$, the latter indicating a strong negative relationship between the two series of innovations (leverage) while maintaining the martingale difference characteristic of $z_t = \exp(x_t/2)\epsilon_t$. Note however that in both cases x_t and u_t are uncorrelated at all leads and lags, satisfying assumption A.4, and the results obtained with the two values of ρ are practically indistinguishable. Therefore, only the results with $\rho = 0$ are shown hereafter; the results with $\rho = -0.8$ are available upon request. Model 2 includes a short memory component in the form of an $AR(1)$ polynomial with a large positive coefficient. This component adds spectral power to the spectral pole caused by the fractional difference operator and makes local estimates of d_0 highly biased. The effect of this bias on the estimation of the signal is analyzed here. The signal in Model 3 contains a pseudo cyclical component such that $f_x(\lambda)$ shows a peak at a frequency close to $\pi/2$. In this case the structure of the signal at frequencies far from the origin is more complex, although knowledge of this complexity is not required at any point in order to implement the signal extraction strategies. The signal in Model 4 is nonstationary but mean reverting and is generated as

$$y_t = y_0 + \kappa \sum_{s=1}^{t} v_s + u_t$$

with $y_0 = \sum_{s=-1000}^{0} v_s$, $(1-L)^{-0.2} v_s = w_s^*$ and the rest of parameters as before. The memory parameter of the signal is now $d = 0.8$. Model 5 is introduced to asses the impact of higher order dependences between signal and noise, while keeping them uncorrelated. The innovations of signal and noise are assumed to be dependent but uncorrelated ARCH processes (for more details see Wong and Li, 1997 and Iglesias and Phillips, 2005). The values of α are selected as in Wong and Li (1997). Considering that

$$Ew_t^2 = a_1 \frac{\alpha_0}{1 - \alpha_2 - \alpha_1} \text{ and } Eu_t^2 = a_2 \frac{\alpha_0}{1 - \alpha_2 - \alpha_1},$$

the constants (a_1, a_2) are chosen to satisfy $E(w_t^2) = 1$ and $E(u_t^2) = \pi^2/2$ such that both signal and noise have the same variances as in Model 1. Note however that, contrary to the previous models, Model 5 does not correspond to any LMSV since the exponential of y_t does not have a martingale difference structure and $\exp(u_t/2)$ is not *i.i.d.*. Finally Model 6 is a FIEGARCH model with the innovations of the signal sharing the same mean zero and unit variance as in Model 1 but with signal and noise correlated. In fact, since $\epsilon_t \sim \mathcal{NID}(0,1)$, $cov(u_{t-1}, w_t) = cov(\log \epsilon_{t-1}^2, g(\epsilon_{t-1}))/\sqrt{var(g(\epsilon_{t-1}))} = 1.23$ and the weights in the optimal filter are

$$\psi_j = \mathbf{1}_{j=0} - \frac{1}{\pi}\int_0^\pi \frac{\theta}{f_y(\lambda)}\cos(j\lambda)\mathrm{d}\lambda - \frac{1}{\pi}\int_0^\pi \frac{f_{ux}^R(\lambda)}{f_y(\lambda)}\cos(j\lambda)\mathrm{d}\lambda + \frac{1}{\pi}\int_0^\pi \frac{f_{ux}^I(\lambda)}{f_y(\lambda)}\sin(j\lambda)\mathrm{d}\lambda$$

where f_{ux}^R and f_{ux}^I are the real and imaginary parts of the cross spectral density function of u_t and x_t. We include this model because of its popularity and as a tool for analyzing the effects of ignoring the correlation between signal and noise in the definition of the filter for signal extraction.

In order to assess the impact of different signal to noise ratios, we consider two different values of κ that give rise to long run noise to signal ratios (NSR hereafter) $f_u(0)/\kappa^2 f_w(0) = \pi^2,\ 5\pi^2$. These NSRs were chosen because they are close to those found empirically when an LMSV model is fitted to financial time series (see Breidt et al. 1998 and Pérez and Ruiz, 2003 among others). The first is close to the ratios considered in Deo and Hurvich (2001) and Sun and Phillips (2003). The second involves a much larger variance of the noise than of the signal. In this case signal estimation is far more difficult. Since the variance of u_t is $\sigma_u^2 = \pi^2/2$ we take $\kappa^2 = 0.5, 0.1$ times $(2\pi f_w(0))^{-1}$. The sample size is $n = 2048$ which is comparable to the size of many financial series, e.g. that analyzed in the next section, and permits the exact use of the Fast Fourier Transform.

Six different estimators of the volatility component are considered:

1. $\hat{x}_{t|n}^{(1)}$ is the frequency domain estimator defined in (10) and (7) in the paper with $M = 100$ (larger values do not result in any improvement) and f and θ estimated by $\hat{f}_y$ and the local Whittle estimator $\hat{\theta}$.

2. $\hat{x}_{t|n}^{(2)}$ is the infeasible frequency domain estimator with $M = 100$ and true f_y and θ.

3. $\hat{x}_{t|n}^{(3)}$ is the proposal by Harvey (1998), $\tilde{x}$ in (1), with true variance and covariances (infeasible). Instead of inverting the 2048×2048 matrix Σ_y we follow the suggestion by Harvey (1998) and consider weights for a smaller sample size. In particular we use weights corresponding to a sample size of 256, padding the rest of the values with zeroes.

4. $\hat{x}_{t|n}^{(4)}$ is a plug-in version of $\hat{x}_{t|n}^{(3)}$ where the covariances of y_t have been replaced by their sample counterparts and σ_u^2 by the local Whittle estimate as in $\hat{x}_{t|n}^{(1)}$.

5. $\hat{x}^{(5)}_{t|n}$ is obtained by smoothing via the Kalman filter based on an $AR(10)$[1] model for the signal and the parameters are estimated by parametric Whittle estimation[2].

6. $\hat{x}^{(6)}_{t|n} = y_t - \hat{\mu}$, which is often used as an approximation to volatility in financial time series.

The estimators $\hat{x}^{(2)}_{t|n}$ and $\hat{x}^{(3)}_{t|n}$ are clearly not feasible whereas $\hat{x}^{(1)}_{t|n}$ and $\hat{x}^{(4)}_{t|n}$ are their plug-in feasible (after bandwidth selection) versions. $\hat{x}^{(5)}_{t|n}$ is calculated by smoothing in the Kalman filter based on a misspecified AR fit to the long memory volatility component and $\hat{x}^{(6)}_{t|n}$ is a naive option that ignores the existence of noise.

The performance of the different signal extraction strategies is assessed by considering two criteria: The Monte Carlo MSE and the correlation between the true x_t and its estimated counterpart. The MSE and the correlation of the infeasible optimal filter $\tilde{x}_{t|\infty}$ defined in (3) and (4) in the text of the paper with μ known are considered as a benchmark. Note that the filter in $\hat{x}^{(2)}_{t|n}$ differs from the optimal one in the truncation to $M = 100$ lags, the estimation of the constant μ and the discretization to obtain the weights. Similarly $\hat{x}^{(3)}_{t|n}$ differs from the optimal filter in the estimation of μ and the truncation to calculate the weights. In Model 6 the differences are even larger because $\hat{x}^{(2)}_{t|n}$ and $\hat{x}^{(3)}_{t|n}$ are based on a misspecified model which ignores the cross spectral density between signal and noise. In the stationary case both MSE and correlation of the optimal filter can be obtained analytically. Taking into account that the spectral density function of $\tilde{x}_{t|\infty}$ is $|f_{xy}(\lambda)|^2/f_y(\lambda)$ and that the covariance between $\tilde{x}_{t|\infty}$ and x_t is equal to the variance of $\tilde{x}_{t|\infty}$, the MSE and the correlation of $\tilde{x}_{t|\infty}$ with x_t, denoted hereafter as MSE_{opt} and $Corr_{opt}$ respectively, can be easily obtained by numerical integration as

$$MSE_{opt} = E[\tilde{x}_{t|\infty} - x_t]^2 = \int_{-\pi}^{\pi} \frac{f_x(\lambda)f_y(\lambda) - |f_{xy}(\lambda)|^2}{f_y(\lambda)} \mathrm{d}\lambda$$

and

$$Corr_{opt} = \left(\frac{\int_{-\pi}^{\pi} \frac{|f_{xy}(\lambda)|^2}{f_y(\lambda)} \mathrm{d}\lambda}{\int_{-\pi}^{\pi} f_x(\lambda)\mathrm{d}\lambda} \right)^{1/2}.$$

which in Models 1-5 becomes

$$MSE_{opt} = E[\tilde{x}_{t|\infty} - x_t]^2 = \int_{-\pi}^{\pi} \frac{f_x(\lambda)f_u(\lambda)}{f_y(\lambda)} \mathrm{d}\lambda$$

[1]We also tried other truncations for the order of the autorregression and found the results to be similar or worse. For example, for Model 1 the MCMSEs defined in (2) obtained with AR(p) for $p = 10, 15, 20, 25$ are 1.204, 1.332, 1.565 and 1.607, justifying the choice of the smaller truncation 10.

[2]The application of the Kalman filter to construct the likelihood function and estimate the parameters is computationally very demanding and inaccurate in large samples with a large number of parameters to be estimated as in this case. Parametric Whittle estimation is much faster and more reliable in this context.

and

$$Corr_{opt} = \left(\frac{\int_{-\pi}^{\pi} \frac{f_x^2(\lambda)}{f_y(\lambda)} \mathrm{d}\lambda}{\int_{-\pi}^{\pi} f_x(\lambda) \mathrm{d}\lambda} \right)^{1/2}.$$

due to uncorrelation between signal and noise. They are shown in Table 1 for Models 1, 2, 3, 5 and 6 (remember that Model 4 is nonstationary) for both NSRs considered. Both MSE and correlation depend inversely on the NSR because the large NSR is associated in our design of the Monte Carlo with a smaller variance of the signal.

Table 1: MSE and correlation with the infeasible optimal signal extractor

		MSE_{opt}	$Corr_{opt}$
Models 1,5	$NSR = \pi^2$	0.541	0.556
	$NSR = 5\pi^2$	0.137	0.357
Model 2	$NSR = \pi^2$	0.172	0.713
	$NSR = 5\pi^2$	0.053	0.488
Model 3	$NSR = \pi^2$	1.313	0.777
	$NSR = 5\pi^2$	0.467	0.544
Model 6	$NSR = \pi^2$	0.353	0.741
	$NSR = 5\pi^2$	0.086	0.671

To obtain comparable measures for different NSRs we standardize the MSE by the variance of the signal (the differenced signal in Model 4) and define the global typified Monte Carlo Mean Square Error as

$$MCMSE(i) = \frac{1}{\sigma^2} \frac{1}{n} \sum_{t=1}^{n} \frac{1}{N} \sum_{k=1}^{N} (\hat{x}_{t,k|n}^{(i)} - x_{t,k})^2 \tag{2}$$

for $i = 1, 2, ..., 6$ corresponding to the different estimators of the signal, where N is the number of replications, $\sigma^2 = \sigma_x^2$ in Models 1, 2, 3, 5 and 6, $\sigma^2 = \sigma_v^2$ in Model 4 and the subindex t, k indicates observation t in the Monte Carlo replication k. Standardizing by σ^2 enables different situations to be compared directly independently of the variance of the signal, which is lower in the case of the larger NSR, such that the differences in MCMSE are attributable only to the NSR.

The performance of $\hat{x}_{t|n}^{(1)}$ depends on the selection of m for local Whittle estimation and m^* for (pseudo) spectral density estimation. The criteria for bandwidth selection proposed in the paper are not automatic and, although they are data-driven, they require the intervention of the researcher. Therefore, it is interesting to analyze how sensitive the estimation of the signal is to the selection of m and m^*. To that end Table 2 shows the MCMSE and the average correlation of $\hat{x}_{t|n}^{(1)}$ with the true signal (in round brackets) in $N = 1000$ replications of Models 2 and 3 with different choices of m and m^*. The results with Models 1, 4 and 5 are similar to those with Model 2 and are thus omitted. Model 6 is not considered because $\hat{x}^{(1)}_{t|n}$ is based in this case on the erroneous uncorrelation assumption A.4. In general $\hat{x}^{(1)}_{t|n}$ is quite robust to the selection of m^* but different m may lead to significantly different results. Table 2 shows that in Model 2 (and also in Models 1,4 and 5) the larger m is, the lower MCMSE is and the higher the correlation is, even though the bias of the estimation of d in that case is quite large. This can be explained by the fact that d only enters $\hat{x}^{(1)}_{t|n}$ via the estimation of the spectral density function by $\hat{f}_y(\lambda_v)$ over the whole band of Fourier frequencies. A positively biased estimate of d implies an excessive damping of the periodogram due to the factor $|\lambda_v + \lambda_j|^{2\hat{d}}$, but this effect is eventually offset by $\lambda_v^{-2\hat{d}}$. However Model 3 shows greater structure in the spectral density at frequencies far from the origin, and a lower m is recommended. The spectral peak around frequency λ_{512} in Model 3 should be avoided in local Whittle estimation and bandwidths containing that frequency and neighboring ones lead in general to worse results. Large bandwidths result in a negative bias on the estimation of d such that $|\lambda_v + \lambda_j|^{2\hat{d}}$ is not sufficient to neutralize the divergent behaviour of the periodogram at frequencies close to the origin. This peak can be easily detected in practice by visual inspection of $\hat{f}_y$ at frequencies sufficiently far from the origin. Based on these considerations we choose $(m, m^*) = (1000, 80)$ for Models 1, 2, 4, 5 and 6 and $(m, m^*) = (300, 60)$ in Model 3.

The constant μ is estimated in Models 1, 2, 3, 5 and 6 by the sample mean. In the nonstationary Model 4 the average of the first 10 initial observations is used, which is $O_p(1)$ under the type I definition of nonstationary long memory used here, and gives better results than using y_1. Harvey's method of signal extraction in the time domain is not directly applicable in Model 4 because the variance is undefined. In this case Harvey (1998) suggests extracting the signal in the differenced series and integrating back to get an estimate of the original signal. We follow this idea and estimate the differenced signal as

$$\widehat{\Delta x} = (I - \sigma_u^2 D \Sigma_{\Delta y}^{-1}) \Delta y \tag{3}$$

with D being a matrix with 2 on the leading diagonal, -1 on the first off-diagonals on either side and 0 on the rest of its elements. Note that μ disappears here due to differencing. We could also have used the prior differencing-integration back strategy proposed by Harvey (1998) in order to avoid estimation of the constant μ in the rest of the models, but this approach needs initial values to be selected in the integrating step and performs worse than working directly with the original series (results available upon request), which is one of the main advantages of the strategies in the frequency domain. $\hat{x}^{(3)}_{t|n}$ is then obtained by integrating back as

$$\hat{x}^{(3)}_{t|n} = \hat{x}^{(3)}_{t-1|n} + \widehat{\Delta x}_t\ , \quad t = 2, ..., n,$$

Table 2: Sensitivity to the choice of m and m^*

	Model 2 $NSR = \pi^2$					
	m=40	m=100	m=300	m=600	m=800	m=1000
$m^* = 40$	5.659	3.634	1.367	0.882	0.793	0.771
	(0.317)	(0.438)	(0.575)	(0.633)	(0.647)	(0.655)
$m^* = 60$	5.603	3.619	1.351	0.862	0.773	0.749
	(0.330)	(0.451)	(0.591)	(0.653)	(0.668)	(0.677)
$m^* = 80$	5.563	3.614	1.348	0.857	0.767	0.744
	(0.339)	(0.458)	(0.599)	(0.663)	(0.678)	(0.687)
$m^* = 100$	5.530	3.614	1.351	0.860	0.770	0.746
	(0.347)	(0.463)	(0.604)	(0.668)	(0.684)	(0.693)
	$NSR = 5\pi^2$					
$m^* = 40$	18.500	15.389	10.408	7.177	6.428	5.987
	(0.193)	(0.236)	(0.281)	(0.302)	(0.309)	(0.310)
$m^* = 60$	18.446	15.372	10.381	7.142	6.391	5.949
	(0.205)	(0.252)	(0.302)	(0.326)	(0.334)	(0.335)
$m^* = 80$	18.452	15.405	10.414	7.168	6.416	5.972
	(0.211)	(0.259)	(0.310)	(0.336)	(0.344)	(0.345)
$m^* = 100$	18.488	15.457	10.475	7.222	6.471	6.024
	(0.213)	(0.262)	(0.313)	(0.339)	(0.347)	(0.348)
	Model 3 $NSR = \pi^2$					
	m=40	m=100	m=300	m=600	m=800	m=1000
$m^* = 40$	1.088	0.767	0.512	0.723	0.876	1.041
	(0.590)	(0.692)	(0.740)	(0.626)	(0.505)	(0.555)
$m^* = 60$	1.072	0.763	0.508	0.707	0.844	1.019
	(0.595)	(0.695)	(0.743)	(0.630)	(0.518)	(0.567)
$m^* = 80$	1.060	0.763	0.509	0.698	0.821	1.004
	(0.598)	(0.695)	(0.742)	(0.629)	(0.525)	(0.574)
$m^* = 100$	1.053	0.765	0.514	0.693	0.805	0.994
	(0.598)	(0.693)	(0.739)	(0.625)	(0.528)	(0.579)
	$NSR = 5\pi^2$					
$m^* = 40$	3.269	2.620	1.890	1.297	1.438	2.325
	(0.353)	(0.398)	(0.431)	(0.293)	(0.363)	(0.384)
$m^* = 60$	3.264	2.622	1.891	1.280	1.430	2.321
	(0.353)	(0.397)	(0.431)	(0.291)	(0.364)	(0.385)
$m^* = 80$	3.271	2.635	1.905	1.279	1.437	2.330
	(0.347)	(0.390)	(0.423)	(0.282)	(0.358)	(0.379)
$m^* = 100$	3.284	2.653	1.925	1.286	1.451	2.343
	(0.340)	(0.380)	(0.411)	(0.269)	(0.348)	(0.370)

MCMSE and correlation with true signal (in round brackets) of $\hat{x}^{(1)}_{t|n}$ with different m and m^*.

with $\hat{x}^{(3)}_{1|n} = 0$. The feasible version $\hat{x}^{(4)}_{t|n}$ is similarly obtained with the local Whittle estimate of σ_u^2 in the original series and the sample autocovariances of the differenced series. Using the original series to estimate σ_u^2 guarantees its consistency. Had we used the differenced series we would have had to deal with an antipersistent signal perturbed by a noninvertible noise, where the consistency of the local Whittle estimator has not been established. The sample autocovariances are however those of the differenced series. Although their statistical properties are unknown when applied to a signal perturbed by a noninvertible noise, their large negative bias in a stationary long memory plus noise context (Pérez, 2000) leads us to conjecture that there will also be a large (even larger) bias in an antipersistent signal plus noninvertible noise series. Frequency domain methods do not suffer from this problem because the ψ_j are fully estimated in the original series.

The top number in each cell in Table 3 shows the MCMSE for every model and signal extractor for $N = 1000$ replications. Note that the MCMSE corresponding to Model 4 is not comparable with the other models since the MSE is standardized by a different quantity (the variance of the differenced signal). The number in the middle, in round brackets, is the global correlation, which is constructed as the average of the sample correlations between the series of true and estimated signals over the 1000 replications. Finally the bottom number in each cell, in square brackets, is the number of times that the Ljung Box statistic does not reject the hypothesis that the first 100 autocorrelations of the squared standardized residuals ($\hat{\varepsilon}^{2(i)}_{t|n} = \exp(y_t - \hat{x}^{(i)}_{t|n})$) are null at the 5% significance level. Table 3 also shows the standardized MSE_{opt} and $Corr_{opt}$ (in italics) in the first column as a benchmark (not available for the nonstationary Model 4).

The performances of the infeasible techniques $\hat{x}^{(2)}_{t|n}$ and $\hat{x}^{(3)}_{t|n}$ are similar in all cases. The MCMSE is larger than optimal in both cases, but the correlation can be larger than that obtained with the optimal filter because optimality has been defined in a MSE sense. The feasible (after bandwidth selection) frequency domain version is significantly better than its time domain counterpart in terms of MSE, correlation with the true signal and nonrejection of no autocorrelation in the squared standardized residuals. The large bias of the sample autocovariances (Hosking, 1996) in an LMSV model (Pérez 2000) helps to explain the worse behavior of the time domain estimates. Considering the feasible strategies in Models 1-5, where the null cross spectral density is correctly imposed in the filters, $\hat{x}^{(1)}_{t|n}$ is only beaten by $\hat{x}^{(5)}_{t|n}$ in terms of MSE in the stationary cases with large NSR. But even in those cases $\hat{x}^{(1)}_{t|n}$ is the best option in terms of correlation with the true signal and no autocorrelation in the squared standardized residuals. The naive $\hat{x}^{(6)}_{t|n}$ is the worst option, and the time domain proposal $\hat{x}^{(4)}_{t|n}$ is the second worst. Regarding the FIEGARCH in Model 6, the unaccounted correlation between signal and noise significantly lowers the performance of the strategies based on a null cross spectral density, as expected. However, the rejection of the absence of autocorrelation in the squared standardized residuals is quite frequent, suggesting that the

Table 3: Global MSE and correlation measures

	$\tilde{x}_{t\|\infty}$	$\hat{x}^{(1)}_{t\|n}$	$\hat{x}^{(2)}_{t\|n}$	$\hat{x}^{(3)}_{t\|n}$	$\hat{x}^{(4)}_{t\|n}$	$\hat{x}^{(5)}_{t\|n}$	$\hat{x}^{(6)}_{t\|n}$
			Model 1				
$NSR=\pi^2$	*0.523*	**0.788**	0.711	0.719	1.299	1.166	4.948
	(0.556)	**(0.545)**	(0.579)	(0.572)	(0.331)	(0.158)	(0.379)
		[842]	[923]	[916]	[422]	[28]	[377]
$NSR=5\pi^2$	*0.662*	4.017	0.871	0.875	6.032	**2.718**	23.983
	(0.357)	**(0.289)**	(0.376)	(0.372)	(0.116)	(0.082)	(0.178)
		[678]	[923]	[918]	[164]	[153]	[375]
			Model 2				
$NSR=\pi^2$	*0.286*	**0.744**	0.638	0.656	1.620	1.616	8.528
	(0.713)	**(0.687)**	(0.734)	(0.716)	(0.389)	(0.083)	(0.271)
		[717]	[869]	[861]	[122]	[0]	[387]
$NSR=5\pi^2$	*0.443*	5.972	0.798	0.803	9.453	**4.257**	41.468
	(0.488)	**(0.345)**	(0.504)	(0.498)	(0.115)	(0.038)	(0.123)
		[650]	[907]	[890]	[116]	[133]	[409]
			Model 3				
$NSR=\pi^2$	*0.368*	**0.508**	0.427	0.428	0.661	0.865	1.439
	(0.777)	**(0.743)**	(0.779)	(0.778)	(0.652)	(0.416)	(0.636)
		[775]	[879]	[872]	[677]	[299]	[413]
$NSR=5\pi^2$	*0.654*	1.891	0.714	0.715	2.444	**1.340**	6.959
	(0.544)	**(0.431)**	(0.545)	(0.544)	(0.297)	(0.206)	(0.345)
		[676]	[925]	[909]	[351]	[47]	[379]
			Model 4				
$NSR=\pi^2$		**86.088**	85.970	85.509	91.806	109.608	94.156
		(0.968)	(0.970)	(0.970)	(0.926)	(0.567)	(0.847)
		[820]	[856]	[899]	[102]	[8]	[321]
$NSR=5\pi^2$		**91.255**	90.235	85.254	124.873	118.380	135.317
		(0.933)	(0.942)	(0.943)	(0.747)	(0.288)	(0.603)
		[692]	[786]	[844]	[33]	[0]	[323]
			Model 5				
$NSR=\pi^2$	*0.523*	**0.811**	0.709	0.709	1.317	1.188	4.944
	(0.556)	**(0.545)**	(0.581)	(0.581)	(0.336)	(0.158)	(0.379)
$NSR=5\pi^2$	*0.662*	4.521	0.865	0.869	6.507	**2.960**	24.010
	(0.357)	**(0.293)**	(0.382)	(0.382)	(0.119)	(0.085)	(0.180)
			Model 6				
$NSR=\pi^2$	*0.341*	3.455	0.671	0.674	3.516	**1.209**	4.965
	(0.741)	**(0.430)**	(0.638)	(0.637)	(0.422)	(0.157)	(0.373)
		[181]	[162]	[167]	[13]	[18]	[398]
$NSR=5\pi^2$	*0.416*	9.784	0.789	0.801	10.586	**2.727**	24.009
	(0.671)	**(0.319)**	(0.516)	(0.500)	(0.235)	(0.029)	(0.176)
		[508]	[160]	[163]	[177]	[3]	[407]

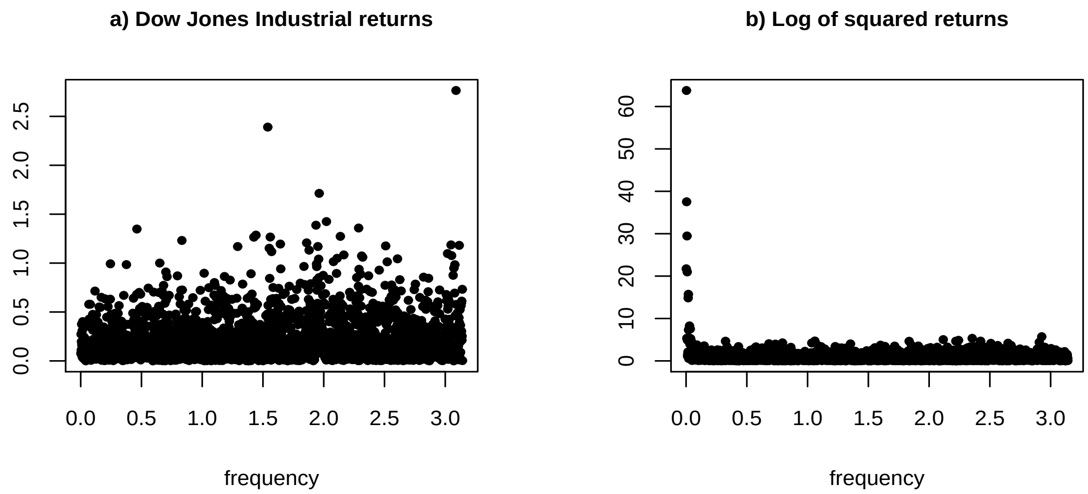
Note: MCMSE, global correlation between x_t and $\hat{x}^{(i)}_{t|n}$ (in round brackets) and nonrejections of no correlation in squared standardized residuals (in square brackets). Optimals in italic (benchmark) and best feasibles in bold.

LMSV model with uncorrelated signal and noise is not suitable for capturing the behavior of these series.

2 Estimating the volatility of daily Dow Jones returns

The returns of the daily Dow Jones Industrial Index, z_t, from December 12, 1996 to November 14, 2012 ($n = 4008$) are analyzed with the proposed signal extraction strategy. Figure 1 shows the periodogram of the returns and of the log of squared centered returns $y_t = \log(z_t - \bar{z})^2$, justifying the lack of linear correlation in the returns and the high persistence in the log of squares, a behavior consistent with LMSV. To corroborate the visual impression of absence of autocorrelation in the returns we use the corrected version of the Box-Pierce statistic as suggested by Deo (2000) and Lobato et al. (2001), which is robust to the presence of higher order dependence typical of financial time series. The corrected Box-Pierce statistic for the first 100 autocorrelations takes a value of 113.75 with a *p-value* of 0.164, confirming the absence of linear correlation in the returns for the usual levels of significance.

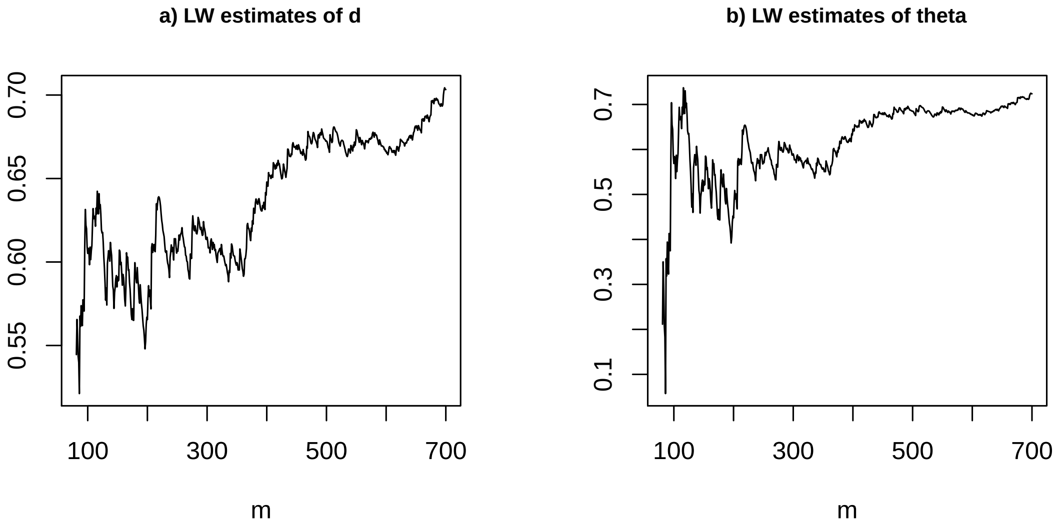
Figure 1: Periodogram: Daily Dow Jones returns (12/12/1996-01/18/2011)



The local Whittle estimates of the memory parameter d and θ in y_t are displayed in Figure 2 for a grid of bandwidths $m = 81, ..., 700$. There is a notable positive correlation between the two series of estimates, which is expected because the asymptotic correlation between the local Whittle estimators of d and θ_{10} is $\sqrt{1+4d}/(1+2d)$, that is between 0.80 and 0.85 for d between 0.5 and 0.75. The local Whittle estimates plugged into the formulae for signal extraction are obtained with $m = 500$, giving $\hat{d} = 0.67$ and $\hat{\theta} = 0.69$. This value of

m is chosen because it falls within a stable range of estimates. Note also that for most of the bandwidths considered in Figure 2 the estimates are spread within a narrow band (between 0.6 and 0.7 for d and 0.5, 0.7 for θ) such that other choices would not significantly alter the results obtained hereafter.

Figure 2: Local Whittle estimates in $y_t = \log(z_t - \bar{z})^2$



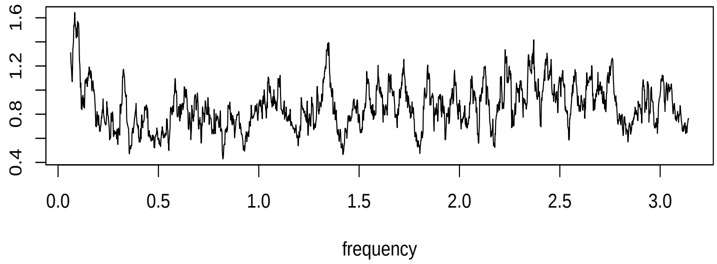
The choice of m^* in step 2 is based on the smoothness of the spectral density at frequencies far from the origin. Figure 3 shows $\hat{f}(\lambda)$ for $m^* = 10$, 60 and 120, where the first 40 Fourier frequencies are omitted to avoid a masking effect of the predominant pole at the origin. The low m^* seems to lead to a very rough estimate but with no significant prevalence of any interval of frequencies. The estimate with $m^* = 60$ seems to reflect some short memory behavior, which is masked with the largest m^*. Based on this reasoning $m^* = 60$ seems a sensible choice. Note also that according to the sensitivity analysis in the previous section neighboring values of m^* are expected to lead to similar results. However, to analyze the sensitivity of the proposed methodology to the choice of the bandwidth in this particular series we consider the three options $m^* = 10$, 60 and 120 in the following steps.

In step 3 we calculate $\hat{\psi}_j$ for the three different m^* considered and chose the truncation point M as the lowest value such that $|\hat{\psi}_j| \leq 0.002$, $\forall j > M$. Figure 4 shows $\hat{\psi}_j$ as a function of j, together with the choice of M, which is 1250, 120 and 45 for $m^* = 10$, 60 and 120 respectively. Finally, since the estimates of d fall well within the nonstationary region we estimate the constant by $\hat{\mu} = \sum_{t=1}^{10} y_t/10$.

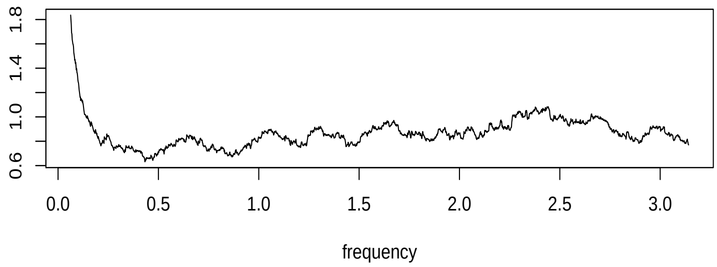
Figure 5 shows the series of returns z_t together with the estimates of the variances of the

Figure 3: Spectral density estimation

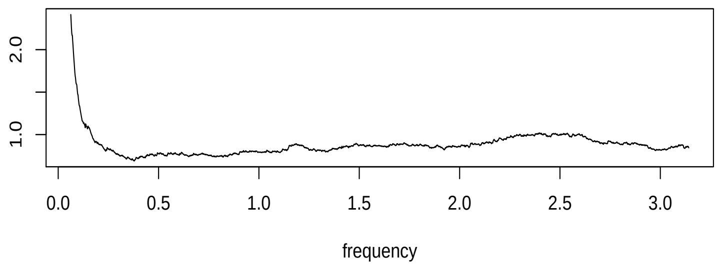
(a) $m^* = 10$



(b) $m^* = 60$



(c) $m^* = 120$



returns conditional on the volatility component in an LMSV model calculated as

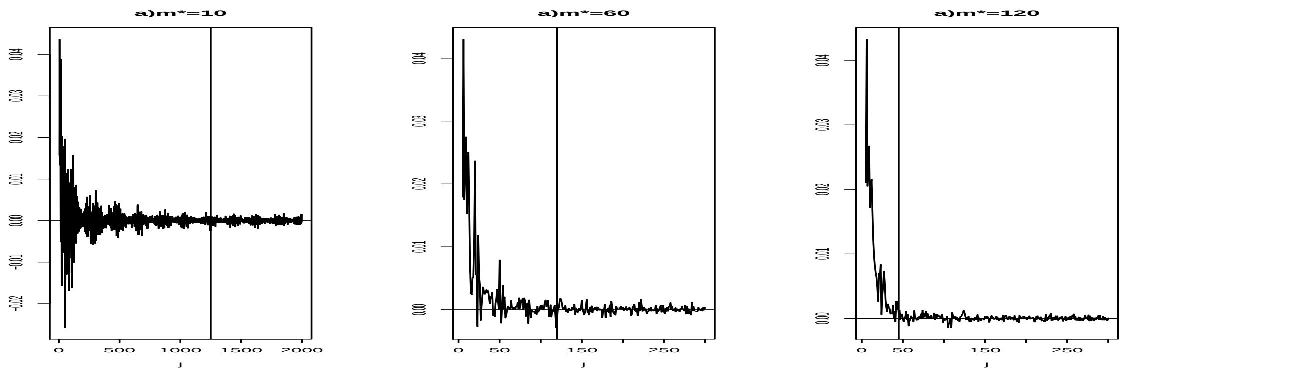
$$\hat{\sigma}_t^2 = \hat{\sigma}^2 \exp(\hat{x}_{t|n}^{(1)})$$

where

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{t=1}^{n} z_t^2 \exp(-\hat{x}_{t|n}^{(1)})$$

as suggested in Harvey (1998), for $m^* = 10$, 60 and 120. All three show similar shapes, especially with the two larger bandwidths, evidencing the low sensitivity of the procedure to

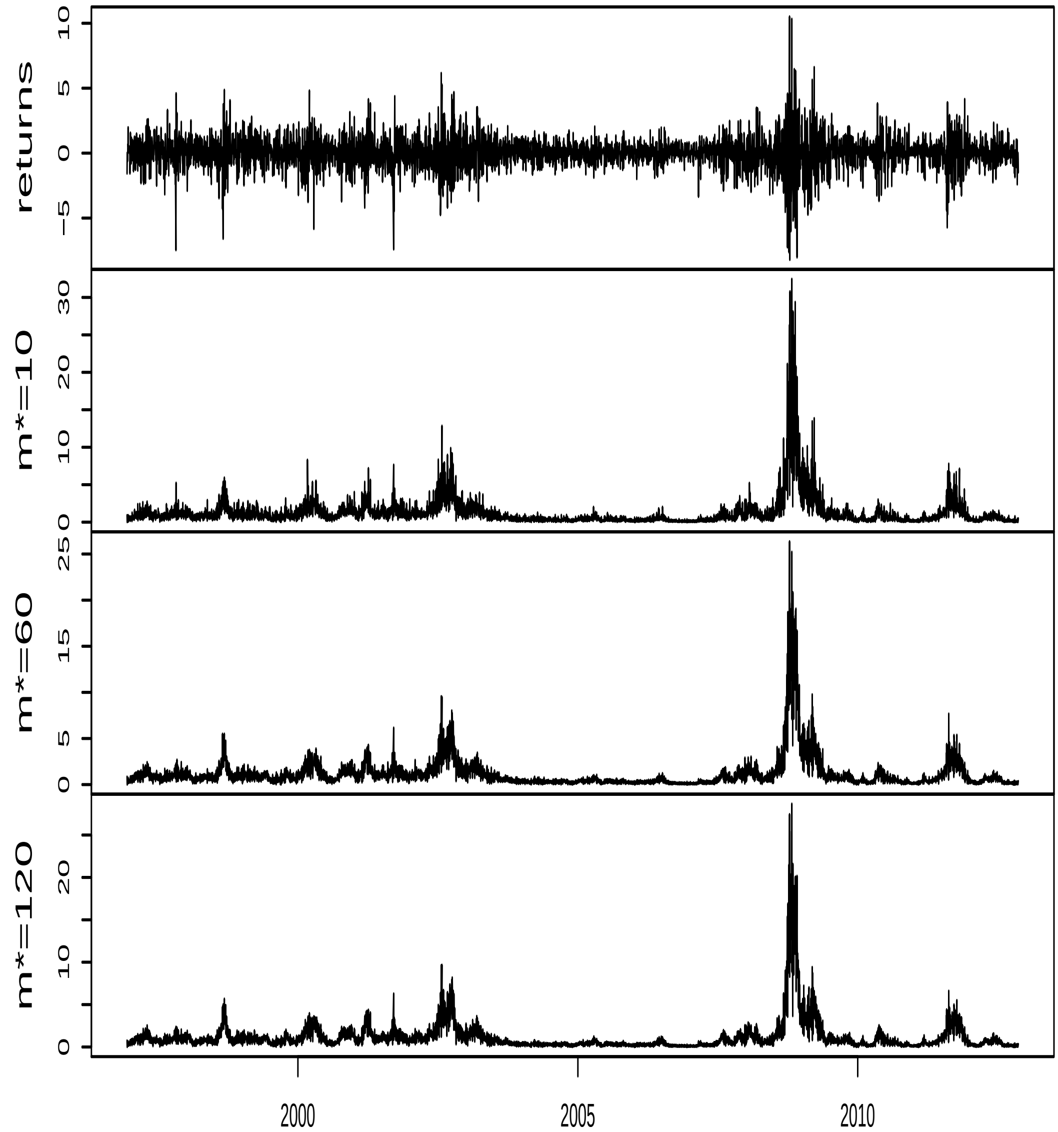
Figure 4: Weights $\hat{\psi}_j$ and M



the choice of m^* (M). The increase in volatility in the subprime crisis dominates the figure, with conditional variances three times larger than the second peak in importance. The increase is significantly steep after September 2008, coinciding with the Federal takeover of Fannie Mae and Freddie Mac and the bankruptcy of Lehman Brothers (mid September 2008), the defeat of the Emergency Economic Stabilization Act in the United States House of Representatives (end of September 2008), the worst week for the stock market in 75 years (second week of October, coinciding with the largest estimated volatility) and the problems of Citigroup with the 60% fall in its share price (November 2008). The second largest peak corresponds to the second half of 2002. This period is post September 11, 2001, and only shows a very short period of high volatility around September 17, the first trading day after 9/11. It is also prior to the beginning of the war in Iraq (March 2003). The months prior to the attack were of great uncertainty. In July 2002 president Bush confirmed a major shift in national security strategy from containment to preemption, increasing uncertainty on the markets. Large volatilities also turn up in October motivated by Congress' authorization of President Bush to use military force against Iraq. The launch of the war however did not increase volatility because it implied a reduction of uncertainty with a general belief that the war was not going to last long. The third highest peak corresponds to the second half of 2011, coinciding with the debt-ceiling crisis in the USA. Other peaks in volatility can also be observed in Figure 5, e.g. at the end of August 1998, coinciding with the Russian crisis, which together with the Asian crisis and the fears of further problems in South America produced a sharp fall in stock markets, and the end of 1999 and the first half of 2000 due to the Argentinian crisis and the Dot Com crash. However these other peaks are less long-lasting and less significant.

To validate the suitability of the volatility estimates we check whether the squared standardized residuals $\hat{\varepsilon}^2_{t|n} = z_t^2/\hat{\sigma}_t^2$ are uncorrelated. The p-values for the Ljung-Box statistics

Figure 5: Estimation of conditional variance



with the first 100 autocorrelations for $m^* = 10$, 60 and 120 are 0.000, 0.113 and 0.138 respectively. This leads us to discard the estimates of the volatility with $m^* = 10$, while the other two options give similarly valid results, reinforcing the validity of an LMSV model for this series and rejecting the possibility of other options with strong persistent volatility (such as FIEGARCH models) or with spurious long memory (such as breaks in the mean).

References

[1] Breidt, F.J., Crato, N. & P. de Lima (1998) The Detection and Estimation of Long Memory in Stochastic Volatility. *Journal of Econometrics* 83, 325-348.

[2] Chan, N.H. & W. Palma (1998) State space modeling of long-memory processes. *The Annals of Statistics* 26, 719-740.

[3] Deo, R.S. (2000) Spectral tests of the martingale hypothesis under conditional heteroscedasticity. *Journal of Econometrics* 99, 291-315.

[4] Deo, R.S. & C.M. Hurvich (2001) On the log periodogram regression estimator of the memory parameter in long memory stochastic volatility models. *Econometric Theory* 17, 686-710.

[5] Harvey, A.C. (1998) Long memory in stochastic volatitility. In: Knight, J., Satchell, S. (Eds.), *Forecasting Volatility in Financial Markets*, Oxford: Butterworth-Haineman, 307-320.

[6] Hosking, J. R. M. (1996) Asymptotic distributions of the sample mean, autocovariances, and autocorrelations of long-memory time series. *Journal of Econometrics* 73, 261-284.

[7] Iglesias, E.M. & G.D.A. Phillips (2005) Bivariate ARCH models: Finite-sample properties of QML estimator and an application to an LM-type test. *Econometric Theory* 21, 1058-1086.

[8] Lobato, I., Nankervis, J.C. & N.E. Savin (2001) Testing for autocorrelation using a modified Box-Pierce Q test. *International Economic Review* 42, 187-205.

[9] Pérez, A. (2000) *Estimación e Identificación de Modelos de Volatilidad Estocástica con Memoria Larga*. Ph.D. of the University of Valladolid, Spain.

[10] Pérez, A. & E. Ruiz (2003). Properties of the sample autocorrelations of nonlinear transformations in Long Memory in Stochastic Volatility models. *Journal of Financial Econometrics* 1, 420-444.

[11] Sun, Y. & P.C.B. Phillips (2003) Nonlinear log-periodogram regression for perturbed fractional processes. *Journal of Econometrics* 115, 355-389.

[12] Wong, H. & W.K. Li (1997) On a multivariate conditional hetersocedastic model. *Biometrika* 84, 111-123.