A discusión

DETECTING LEVEL SHIFTS IN THE PRESENCE OF CONDITIONAL HETEROSCEDASTICITY*

M. Angeles Carnero, Daniel Peña and Esther Ruiz**

WP-AD 2004-06

Corresponding author: M. Angeles Carnero, Universidad de Alicante, Dpt. Fundamentos del Análisis Económico, Spain. Tel.: +34 965903400 Ext. 3263 / Fax: +34 965903898 / E-mail: acarnero@merlin.fae.ua.es.

Editor: Instituto Valenciano de Investigaciones Económicas, S.A.
Primera Edición Febrero 2004.
Depósito Legal: V-1057-2004

IVIE working papers offer in advance the results of economic research under way in order to encourage a discussion process before sending them to scientific journals for their final publication.

---

* Financial support from projects BEC2000-0167 and BEC2002-03720 by the Spanish Government is acknowledged. Part of this work was carried out while the first author was visiting Nuffield College during the summer 2003. She is indebted to Neil Shephard for financial support and very useful discussions. We are also grateful to Ron Bewley, Juan Carlos Escanciano, Oscar Martinez, Pilar Poncela, Marco Reale and participants of the Fifth Time Series Workshop at Arrábida, Portugal and seminar participants at University of Canterbury and University New South Wales for helpful comments and suggestions. Any remaining errors are our own.

** M.A. Carnero: Dpt. Fundamentos del Análisis Económico, Universidad de Alicante. D. Peña y E. Riuz: Dpt. Estadística y Econometría, Universidad Carlos III de Madrid.

DETECTING LEVEL SHIFTS IN THE PRESENCE OF CONDITIONAL HETEROSCEDASTICITY

M. Angeles Carnero, Daniel Peña and Esther Ruiz

ABSTRACT

The objective of this paper is to analyze the finite sample performance of two variants of the likelihood ratio test for detecting a level shift in uncorrelated conditionally heteroscedastic time series. We show that the behavior of the likelihood ratio test is not appropriate in this context whereas if the test statistic is appropriately standardized, it works better. We also compare two alternative procedures for testing for several level shifts. The results are illustrated by analyzing daily returns of exchange rates.

Key words: EGARCH, GARCH, Likelihood Ratio, Stochastic Volatility.

1 Introduction

It is well known that high frequency financial time series are often characterized by being uncorrelated and conditionally heteroscedastic; see, for example, Bollerslev *et al.* (1994), Ghysels *et al.* (1996) and Shephard (1996) among many others. In empirical studies of these series using several years of observations, we often find level shifts which may be caused by wars, financial crisis, policy interventions, etc. There is a vast literature on testing for level shifts in time series; see, for example, Hawkings (1977), Worsley (1986), Tsay (1988), Krämer *et al.* (1988), Andrews (1993), Balke (1993) and Bai (1994) among many others. One of the main problems when testing for a level shift with unknown change point, τ, is that τ only appears under the alternative hypothesis and not under the null. Consequently, usual tests like, for example, the Likelihood Ratio (LR) test, do not have standard asymptotic distributions even under ideal assumptions on the properties of the series analyzed. Although there are variants of the LR test with well defined asymptotic distributions, their properties in the presence of conditional heteroscedasticity are still unknown. The objective of this paper is to analyze the performance of these tests for detecting a level shift at an unknown point in uncorrelated conditionally heteroscedastic series.

The paper is organized as follows. Section 2 describes several variants of the LR statistic proposed to test for a level shift in uncorrelated time series. In Section 3, we analyze the finite sample size and power of these tests in conditionally heteroscedastic series. In particular, we consider series generated by the following models: Generalized Autoregressive Conditionally

Heteroscedastic (GARCH), Exponential GARCH (EGARCH) and Autoregressive Stochastic Volatility (ARSV). In Section 4, we extend the analysis to the case when there are several level shifts in the series of interest and compare two alternative procedures often used in this case. Section 5 contains an empirical application where a series of daily US Dollar/Spanish Peseta exchange rate returns is analyzed. Finally, Section 6 concludes the paper.

2 Likelihood Ratio tests for level shifts

Consider the series of interest given by

$$y_t = \mu + a_t, \quad t = 1, ..., T \tag{1}$$

where μ is the mean and a_t is an uncorrelated white noise process with zero mean and finite variance, σ^2. If there is a level shift at time τ, the observed series, z_t, is given by

$$z_t = y_t + \omega I(t \geq \tau) \tag{2}$$

where w is the size of the shift and $I(t \geq \tau)$ is the indicator function. We are interested in testing the null hypothesis of no level shifts in the series, i.e. $H_0 : w = 0$ against the alternative $H_1 : w \neq 0$. The LR statistic can be derived from the t-statistic of the Ordinary Least Squares (OLS) estimator of the parameter ω in the following regression:

$$z_t = \mu + \omega I(t \geq m) + a_t, \quad t = 1, ..., T, \quad m = 2, ..., T \tag{3}$$

obtained substituting (1) in (2) where, given that τ is unknown, the change point, m, can occur at any moment between $t = 2$ to T.

The t-statistic of the OLS estimator of ω is given by

$$\lambda_m = \frac{\sum_{t=m}^{T}(z_t - \bar{z})}{\hat{\sigma}_m \sqrt{\frac{(m-1)(T-m+1)}{T}}}, \qquad m = 2, \ldots, T. \tag{4}$$

where $\bar{z} = \sum_{t=1}^{T} z_t/T$ and $\hat{\sigma}_m^2 = \widehat{\sigma}_z^2 - \frac{1}{(m-1)(T-m+1)}\left[\sum_{t=m}^{T}(z_t - \bar{z})\right]^2$ and $\widehat{\sigma}_z^2$ is the sample variance of z_t. From (4), it is easy to obtain the following alternative expression of λ_m

$$\lambda_m = \frac{\bar{z}_2 - \bar{z}_1}{\sqrt{\frac{(m-1)\widehat{\sigma}_1^2 + (T-m+1)\widehat{\sigma}_2^2}{(m-1)(T-m+1)}}}, \qquad m = 2, \ldots, T, \tag{5}$$

where $\bar{z}_1 = \sum_{t=1}^{m-1} z_t/(m-1)$ and $\bar{z}_2 = \sum_{t=m}^{T} z_t/(T-m+1)$ are the sample means before and after time m respectively and $\widehat{\sigma}_1^2$ and $\widehat{\sigma}_2^2$ are the corresponding sample variances. Therefore, λ_m can be interpreted as the t-statistic for the difference of the sample means of the first $m-1$ and the last $T-m+1$ observations.

The LR statistic is given by $\lambda = \max_{m=2,\ldots,T} |\lambda_m|$. If λ is greater than the chosen critical value, then $t^\star$ such that $\lambda = |\lambda_{t^\star}| = \max_{m=2,\ldots,T} |\lambda_m|$, is identified as the instant of the change; see, for example, Hawking (1977), Tsay (1988) and Andrews (1993). If a_t is a Gaussian process, λ_m is, under the null hypothesis, N$(0,1)$, for all m. However, λ diverges asymptotically to infinity; see Hawking (1977) and Andrews (1993) for heuristic and analytic proofs respectively. Figure 1, illustrates this point plotting kernel estimates[1] of the densities of λ computed from 10000 replicates of Gaussian series with zero mean and variance one and $T = 25, 200$ and 15000. Consequently, Tsay (1988) obtained the critical values for λ based on Monte Carlo experiments

[1] The kernel estimates of the densities have been obtained using S-plus 4.5 with a Normal kernel.

and proposed using $3.5, 3$ or 2.5 as critical values. Table 1, that reports empirical percentiles of λ based on the same simulated series as before, shows that, for large enough samples, these critical values correspond approximately to sizes of 5%, 15% and 25% respectively. However, notice that the critical values strongly depend on the sample size of the series analyzed.

Andrews (1993) shows that when the change point is bounded away from both extremes of the sample, λ converges in distribution to a function of the supremum of a Brownian bridge that depends on the proportion of observations discarded on both extremes of the sample; see also Bai (1994). Andrews (1993) suggests to discard 15% of the observations in each extreme and consider $m = [0.15T], ..., [0.85T]$ where $[\cdot]$ is the integer-valued function. However, the main disadvantage of this alternative to obtain a well defined asymptotic distribution is that different tables should be used depending on the particular proportion of the sample discarded.

On the other hand, Bai (1994) shows that using an alternative standardization of the difference between means in (5), it is possible to obtain a statistic that converges asymptotically in distribution[2]. In particular, consider the statistic $\mathbf{e} = \max_{m=2,...,T} |e_m|$, where

$$e_m = \frac{(T - m + 1)(m - 1)(\overline{z}_1 - \overline{z}_2)}{\widehat{\sigma}_z T^{3/2}}. \tag{6}$$

Comparing the λ_m and e_m statistics in expressions (5) and (6) respectively, it is possible to derive the following relationship among them

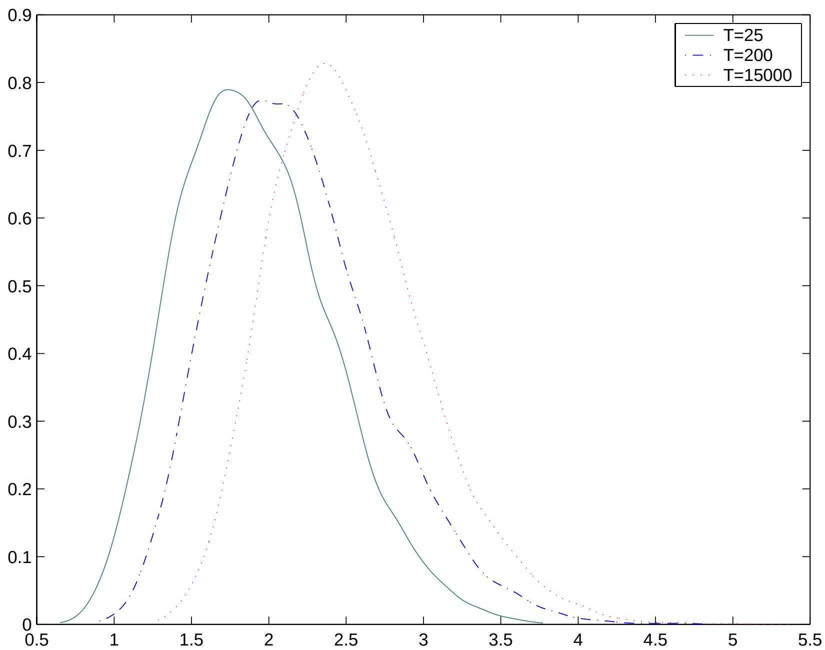
$$e_m = -\frac{\sqrt{(m-1)(T-m+1)}}{T}\frac{\hat{\sigma}_m}{\hat{\sigma}_z}\lambda_m, \qquad m = 2, \ldots, T.$$

[2]A similar test has been proposed by Inclan and Tiao (1994) who use cumulative sums (CUSUM) of squares to detect changes in the variance of independent processes.

Table 1: Percentiles of the empirical distribution of the likelihood ratio test statistic, λ, for Gaussian white noise processes

Percentil	T=25	T=200	T=500	T=1000	T=5000	T=15000
80%	2.62	2.65	2.73	2.77	2.88	2.91
85%	2.78	2.79	2.84	2.89	3.01	3.03
90%	2.98	2.97	3.01	3.04	3.15	3.18
95%	3.36	3.23	3.26	3.28	3.39	**3.43**
99%	4.12	3.77	3.78	3.77	3.87	3.90

Figure 1: Kernel estimates of the density of λ for different sample sizes



Observe that $\hat{\sigma}_m < \hat{\sigma}_z$, and therefore $\frac{\sqrt{(m-1)(T-m+1)}}{T}\frac{\hat{\sigma}_m}{\hat{\sigma}_z} < 1$. Consequently, λ is always greater than $\mathbf{e}$.

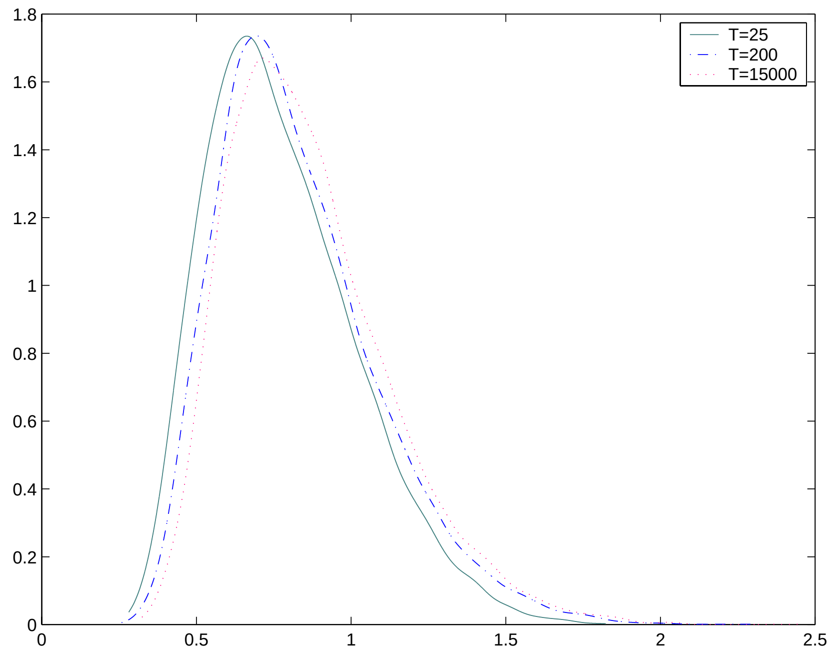
Bai (1994) shows that, under the null hypothesis, the asymptotic distribution of $\mathbf{e}$ is given by the distribution of the supremum of a Brownian bridge, i.e. $P(\mathbf{e} \leq x) \longrightarrow G(x)$ where

$$G(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - 2\sum_{k=1}^{\infty}(-1)^{k+1}e^{-2k^2x^2} & \text{if } x \geq 0 \end{cases}. \quad (7)$$

Table 2 contains percentiles of the asymptotic and empirical distributions of the statistic $\mathbf{e}$, computed from 10000 replicates of Gaussian white noise process with zero mean and variance one for the same sample sizes considered before. Figure 2, which plots kernel estimates of the densities of $\mathbf{e}$, illustrates that the asymptotic distribution is an adequate approximation to the finite sample distribution even for moderately small sample sizes like, for example, $T = 200$.

The power of the λ and $\mathbf{e}$ tests to detect a level shift in uncorrelated homoscedastic series has been analyzed when the size is 5% and 5000 replicates are generated by model (3) with $\mu = 0$ and a_t a Gaussian process with zero mean and variance one. We consider different sizes and moments of the level shift, ω and τ respectively. In particular, we consider increments of ω of 0.2 from 0 to 1 and $\tau = T/10$, $T/2$ and $9T/10$ to analyze the differences on the power when the change occurs at the beginning, the middle or the end of the sample respectively. Finally, the sample sizes considered are $T = 500$, 1000 and 5000. Table 3 reports the results of the corresponding Monte Carlo experiments for λ. As expected, the power increases with the sample size and with the size of the shift and it is higher when the shift occurs in the

Figure 2: Kernel estimates of the density of $\mathbf{e}$ for different sample sizes



middle of the sample than when it happens at the beginning or the end. In any case, for the sample sizes considered, the power is rather high even when ω is relatively small. For example, if $\omega = 0.2$ and $T = 5000$, the power is approximately 0.9 in the extremes and 1 in the middle. Even when $T = 1000$, the power is 0.83 in the extremes and 1 in the center if $\omega = 0.4$. Notice that although these sample sizes could seem too large, they are rather usual when analyzing high frequency financial series.

Table 4 reports, the percentage of rejections of the null hypothesis obtained using the **e** test when the series are generated by the same models as above. Comparing Tables 3 and 4, it is possible to observe that the power of λ is higher when the change occurs at the extremes of the sample while the power of **e** is higher in the middle. On the other hand, the relationship of the latter test with respect the sample size and the size and moment of the change is the same as observed for the LR test.

3 Level shifts and conditional heteroscedasticity

Conditionally heteroscedastic time series are often characterized by being uncorrelated although non-independent. The dynamic evolution of the conditional variances generates autocorrelations of non-linear transformations of absolute observations and non-Gaussian marginal distributions. In this section, we analyze whether the presence of conditional heteroscedasticity affects the performance of the λ and **e** tests described in the previous section.

The Monte Carlo results reported in this section are based on series generated by four different conditionally heteroscedastic models:

Table 2: Percentiles of the empirical distribution of the statistic **e** for Gaussian white noise processes

Percentil	T=25	T=200	T=500	T=1000	T=5000	T=15000	Asymptotic
80%	0.98	1.03	1.05	1.05	1.07	1.07	1.07
85%	1.04	1.10	1.12	1.12	1.13	1.13	1.14
90%	1.11	1.19	1.20	1.21	1.22	1.22	1.22
95%	1.23	1.32	1.34	1.34	1.35	**1.36**	**1.36**
99%	1.44	1.60	1.62	1.62	1.64	1.64	1.63

Table 3: Power of the likelihood ratio test for Gaussian white noise processes

	$\tau = T/10$			$\tau = T/2$			$\tau = 9T/10$		
ω	T=500	T=1000	T=5000	T=500	T=1000	T=5000	T=500	T=1000	T=5000
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.2	0.12	0.24	0.91	0.35	0.66	1.00	0.14	0.23	0.90
0.4	0.47	0.83	1.00	0.94	1.00	1.00	0.49	0.83	1.00
0.6	0.88	1.00	1.00	1.00	1.00	1.00	0.88	1.00	1.00
0.8	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00

Table 4: Power of the **e** test for Gaussian white noise processes

	$\tau = T/10$			$\tau = T/2$			$\tau = 9T/10$		
ω	T=500	T=1000	T=5000	T=500	T=1000	T=5000	T=500	T=1000	T=5000
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.2	0.08	0.12	0.69	0.51	0.82	1.00	0.09	0.14	0.70
0.4	0.24	0.55	1.00	0.98	1.00	1.00	0.26	0.57	1.00
0.6	0.57	0.95	1.00	1.00	1.00	1.00	0.63	0.96	1.00
0.8	0.90	1.00	1.00	1.00	1.00	1.00	0.93	1.00	1.00
1	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00

(i) GARCH(1,1) models with Gaussian errors, given by

$$\begin{aligned} y_t &= \varepsilon_t \sigma_t \\ \sigma_t^2 &= \alpha_0 + \alpha_1 y_{t-1}^2 + \beta \sigma_{t-1}^2 \end{aligned} \tag{8}$$

where ε_t is a Gaussian white noise with zero mean and variance one.

(ii) GARCH(1,1) models defined as in (10) with ε_t having a Student-t with 7 degrees of freedom distribution standardized to have variance one.

(iii) EGARCH(1,1) models with Gaussian errors where y_t is generated as in (10) but the conditional variance is given by

$$\log(\sigma_t^2) = \alpha_0 + \beta \log(\sigma_{t-1}^2) + \alpha_1 \left| \varepsilon_{t-1} - E \left| \varepsilon_{t-1} \right| \right| + \gamma \varepsilon_{t-1}$$

(iv) ARSV(1) models with Gaussian errors given by

$$\begin{aligned} y_t &= \sigma_* \varepsilon_t \sigma_t \\ \log(\sigma_t^2) &= \phi \log(\sigma_{t-1}^2) + \eta_t \end{aligned}$$

where η_t is a Gaussian white noise process with zero mean and variance σ_η^2 independently distributed of ε_t.

The description of these models and their properties can be found, for example, in Carnero *et al.* (2001). The values of the parameters of the previous models, reported in Table 5, have been chosen to represent the parameters usually estimated with real time series of financial returns. We have considered three alternative sample sizes: $T = 500$, 1000 and 5000. The number of replicates when analyzing the size of the tests is 10000 while we generate 5000 replicates to study the power. All the experiments have been carried out in a Pentium III computer using our own Fortran codes.

Table 5: Empirical sizes of the λ and $\mathbf{e}$ tests in conditional heteroscedastic and leptokurtic models when the nominal size is 5%

			$T = 500$		$T = 1000$		$T = 5000$	
Model	Parameters	Kurtosis	λ	$\mathbf{e}$	λ	$\mathbf{e}$	λ	$\mathbf{e}$
	(0.10,0.10,0.80)	3.35	0.05	0.04	0.05	0.04	0.07	0.05
GARCH	(0.02,0.10,0.88)	6.06	0.08	0.05	0.09	0.05	0.10	0.05
$(\alpha_0, \alpha_1, \beta)$	(0.07,0.05,0.88)	3.11	0.04	0.04	0.04	0.04	0.06	0.05
	(0.05,0.15,0.80)	5.57	0.08	0.05	0.09	0.05	0.11	0.05
	(0.10,0.10,0.80)	6.33	0.08	0.05	0.09	0.05	0.11	0.05
GARCH-t_7	(0.02,0.10,0.88)	$\nexists$	0.12	0.06	0.13	0.05	0.15	0.05
$(\alpha_0, \alpha_1, \beta)$	(0.07,0.05,0.88)	5.40	0.06	0.05	0.07	0.05	0.08	0.05
	(0.05,0.15,0.80)	65	0.12	0.05	0.13	0.05	0.15	0.05
	(-0.004,0.20,0.95,0.05)	3.66	0.06	0.05	0.07	0.04	0.08	0.05
EGARCH	(-0.001,0.10,0.98,-0.05)	3.56	0.06	0.05	0.06	0.04	0.08	0.05
$(\alpha_0, \alpha_1, \beta, \gamma)$	(-0.004,0.15,0.95,0.10)	3.74	0.06	0.05	0.07	0.05	0.08	0.05
	(-0.010,0.30,0.98,-0.10)	11.47	0.15	0.06	0.16	0.05	0.17	0.06
	(0.77,0.90,0.10)	5.08	0.10	0.04	0.11	0.04	0.12	0.05
ARSV	(0.78,0.95,0.05)	5.01	0.12	0.05	0.14	0.05	0.15	0.05
$(\sigma^2_\star, \phi, \sigma^2_\eta)$	(0.29,0.98,0.10)	37.48	0.20	0.06	0.20	0.05	0.20	0.05
	(0.80,0.98,0.02)	4.97	0.14	0.05	0.16	0.05	0.18	0.05
Student-t_ν	$\nu = 5$	9	0.05	0.04	0.06	0.04	0.07	0.04
Student-t_ν	$\nu = 7$	5	0.04	0.04	0.04	0.04	0.06	0.05

We analyze first whether the presence of the types of conditional heteroscedasticity considered in this paper, affects the size of the λ and $\mathbf{e}$ tests. Table 5 reports their empirical sizes when the nominal size is 5%[3]. The critical value for λ has been taken from Table 1 as the value corresponding to the sample distribution when $T = 15000$, i.e. 3.43. The critical value for $\mathbf{e}$ has been taken from its asymptotic distribution. Looking at the results for the λ test, the first conclusion is that, with one exception, the empirical size is greater than the nominal. This result seems surprising looking at the results in Table 1 for the Gaussian series, because we are using a critical value larger than the values corresponding to the sample sizes considered in Table 5. Therefore, a smaller size should be expected. In some cases, the size distortions are huge; see, for example, the ARSV(1) models, where some of the empirical sizes are double than the nominal. It seems that, the size distortions of λ are larger the larger the kurtosis of y_t. For models with similar kurtosis, the distortions are larger in the more persistent cases. Furthermore, the gap between the nominal and empirical sizes increases with the sample size.

To analyze whether these size distortions of the λ test are attributable only to the conditional heteroscedasticity or they are the result of the lack of Gaussianity of the GARCH and ARSV models, Table 5 also reports the sizes of λ when the series are generated by homoscedastic although leptokurtic white noises. In particular, we generate series by two Student-t distributions with 5 and 7 degrees of freedom. It can be observed that, in these cases, the size of λ is close to the nominal. Therefore, it is possible to conclude that the

[3]Results for alternative nominal sizes are available from the authors upon request.

size distortions observed before are mainly due to the presence of conditional heteroscedasticity.

Looking at the size results of the **e** test, we can observe that, for all the models and sample sizes considered, the empirical size is very close to the nominal. Consequently, the results reported in Table 5 suggest that when the series are GARCH or ARSV, the size of the **e** test is not affected.

Now, we study the power of the λ and **e** tests when the series are conditionally heteroscedastic. Table 6 shows the empirical powers of the LR test, λ, when there is a level shift for some selected conditionally heteroscedastic models. In particular, we consider two GARCH(1,1) models, with Gaussian and Student-t innovations respectively, and parameters $(\alpha_0, \alpha_1, \beta) = (0.02, 0.10, 0.88)$ in both cases. We also consider an EGARCH model with parameters $(\alpha_0, \alpha_1, \beta, \gamma) = (-0.001, 0.10, 0.98, -0.05)$ and, finally, an ARSV model with parameters $(\sigma_\star^2, \phi, \sigma_\eta^2) = (0.8, 0.98, 0.02)$. The design for the level shift is the same considered in the previous section.

Comparing Tables 3 and 6, it is possible to observe that for all, ω, τ and T, the power of λ decreases when the series are generated by conditionally heteroscedastic models with respect to the powers obtained in Gaussian white noise series. Notice that, in some cases, the lost of power can be very important. For example, when $\omega = 0.2$, $\tau = T/2$ and $T = 1000$, the power is 0.66 when the series is Gaussian while, if the series is GARCH, the powers are 0.51 and 0.37 depending on whether the innovations are Gaussian or Student-t. On the other hand, if the series is EGARCH the power is 0.56 while if it is ARSV is 0.31, less than half than in the conditionally homoscedastic Gaussian model.

Table 6: Power of the likelihood ratio test for conditionally heteroscedastic time series

Gaussian GARCH(1,1) with parameters (0.02,0.10,0.88)									
	$\tau = T/10$			$\tau = T/2$			$\tau = 9T/10$		
ω	T=500	T=1000	T=5000	T=500	T=1000	T=5000	T=500	T=1000	T=5000
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.2	0.10	0.16	0.76	0.25	0.51	1.00	0.09	0.15	0.77
0.4	0.35	0.69	1.00	0.87	0.99	1.00	0.37	0.69	1.00
0.6	0.76	0.97	1.00	0.99	1.00	1.00	0.79	0.97	1.00
0.8	0.95	1.00	1.00	1.00	1.00	1.00	0.96	1.00	1.00
1	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
GARCH(1,1)-t_7 with parameters (0.02,0.10,0.88)									
	$\tau = T/10$			$\tau = T/2$			$\tau = 9T/10$		
ω	T=500	T=1000	T=5000	T=500	T=1000	T=5000	T=500	T=1000	T=5000
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.2	0.07	0.11	0.60	0.16	0.37	0.99	0.07	0.10	0.60
0.4	0.26	0.56	1.00	0.78	0.97	1.00	0.25	0.56	1.00
0.6	0.65	0.93	1.00	0.97	1.00	1.00	0.67	0.94	1.00
0.8	0.89	0.99	1.00	1.00	1.00	1.00	0.91	0.99	1.00
1	0.97	1.00	1.00	1.00	1.00	1.00	0.97	1.00	1.00
Gaussian EGARCH(1,1) with parameters (-0.001,0.10,0.98,-0.05)									
	$\tau = T/10$			$\tau = T/2$			$\tau = 9T/10$		
ω	T=500	T=1000	T=5000	T=500	T=1000	T=5000	T=500	T=1000	T=5000
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.2	0.12	0.19	0.84	0.26	0.56	1.00	0.09	0.16	0.85
0.4	0.38	0.72	1.00	0.90	1.00	1.00	0.38	0.76	1.00
0.6	0.79	0.99	1.00	1.00	1.00	1.00	0.82	0.99	1.00
0.8	0.97	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00
1	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00	1.00
Gaussian ARSV(1) with parameters (0.8,0.98,0.02)									
	$\tau = T/10$			$\tau = T/2$			$\tau = 9T/10$		
ω	T=500	T=1000	T=5000	T=500	T=1000	T=5000	T=500	T=1000	T=5000
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.2	0.09	0.12	0.64	0.13	0.31	1.00	0.07	0.09	0.66
0.4	0.26	0.50	1.00	0.69	0.98	1.00	0.18	0.47	1.00
0.6	0.58	0.90	1.00	0.98	1.00	1.00	0.55	0.94	1.00
0.8	0.85	1.00	1.00	1.00	1.00	1.00	0.88	1.00	1.00
1	0.97	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00

To show that the reduction of power of the λ test in the presence of conditional heteroscedasticity is not only attributable to the models considered in Table 6, Table 7 reports the corresponding powers for all the models considered when analyzing the size. It is rather obvious that the power of λ decreases for all the models considered and that the reduction in power is larger the larger the kurtosis of y_t. It is also interesting to notice that the behavior of the power is similar between the GARCH-t and ARSV models and between the GARCH-N and EGARCH models, being much smaller in the former than in the latter. This result is consistent with the results in Carnero *et al.* (2003) who show that the statistical properties of the first two and the last two models are similar. Finally, Table 7 also reports the powers of the λ test when the series are generated by homoscedastic Student-t white noises. Notice that, in these cases, the loss of power is rather small. Consequently, the problems of the λ test to detect level shifts can be attributable to the dynamic evolution of the conditional variance.

Table 8 reports the powers of the $\mathbf{e}$ test when artificial series are generated by the same models as in Table 6. Comparing Tables 4 and 8, it is possible to observe that the power of the $\mathbf{e}$ test is only marginally affected by the presence of conditional heteroscedasticity. Given that in these circumstances, as we have seen before, the power of λ decreases, there is an important increase in the power of $\mathbf{e}$ with respect to λ, specially when the change happens in the middle of the sample. Consider, for example, the series generated by the GARCH-t_7 model with $T = 1000$ and $\omega = 0.2$. If the level shift occurs at $\tau = T/2$, the powers of λ and $\mathbf{e}$ are 0.37 and 0.82 respectively. Table 7 shows that, for this particular case, the power of the $\mathbf{e}$ test is not affected by the

Table 7: Power of the λ and $\mathbf{e}$ tests for different conditional heteroscedastic and leptokurtic models when $T = 1000$ and the level shift occurs at $\tau = T/2$ with a size of 0.2

Model	Parameters	Kurtosis	λ	$\mathbf{e}$
	(0.10,0.10,0.80)	3.35	0.58	0.82
GARCH	(0.02,0.10,0.88)	6.06	0.51	0.81
$(\alpha_0, \alpha_1, \beta)$	(0.07,0.05,0.88)	3.11	0.62	0.82
	(0.05,0.15,0.80)	5.57	0.48	0.82
	(0.10,0.10,0.80)	6.33	0.47	0.82
GARCH-t_7	(0.02,0.10,0.88)	$\nexists$	0.37	0.82
$(\alpha_0, \alpha_1, \beta)$	(0.07,0.05,0.88)	5.40	0.52	0.82
	(0.05,0.15,0.80)	65	0.36	0.83
	(-0.004,0.20,0.95,0.05)	3.66	0.53	0.82
EGARCH	(-0.001,0.10,0.98,-0.05)	3.56	0.56	0.82
$(\alpha_0, \alpha_1, \beta, \gamma)$	(-0.004,0.15,0.95,0.10)	3.74	0.47	0.81
	(-0.010,0.30,0.98,-0.10)	11.47	0.34	0.81
	(0.77,0.90,0.10)	5.08	0.44	0.82
ARSV	(0.78,0.95,0.05)	5.01	0.36	0.81
$(\sigma_\star^2, \phi, \sigma_\eta^2)$	(0.29,0.98,0.10)	37.48	0.29	0.77
	(0.80,0.98,0.02)	4.97	0.31	0.78
Student-t_ν	$\nu = 5$	9	0.61	0.83
Student-t_ν	$\nu = 7$	5	0.61	0.82

Table 8: Power of the **e** test for conditionally heteroscedastic time series

Gaussian GARCH(1,1) with parameters (0.02,0.10,0.88)									
	$\tau = T/10$			$\tau = T/2$			$\tau = 9T/10$		
ω	T=500	T=1000	T=5000	T=500	T=1000	T=5000	T=500	T=1000	T=5000
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.2	0.09	0.14	0.68	0.53	0.81	1.00	0.09	0.14	0.69
0.4	0.28	0.56	1.00	0.97	1.00	1.00	0.29	0.57	1.00
0.6	0.63	0.93	1.00	1.00	1.00	1.00	0.66	0.94	1.00
0.8	0.89	0.99	1.00	1.00	1.00	1.00	0.91	0.99	1.00
1	0.97	1.00	1.00	1.00	1.00	1.00	0.98	1.00	1.00
GARCH(1,1)-t_7 with parameters (0.02,0.10,0.88)									
	$\tau = T/10$			$\tau = T/2$			$\tau = 9T/10$		
ω	T=500	T=1000	T=5000	T=500	T=1000	T=5000	T=500	T=1000	T=5000
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.2	0.10	0.15	0.69	0.55	0.82	1.00	0.10	0.15	0.69
0.4	0.29	0.58	1.00	0.96	1.00	1.00	0.29	0.59	1.00
0.6	0.64	0.93	1.00	1.00	1.00	1.00	0.66	0.93	1.00
0.8	0.88	0.99	1.00	1.00	1.00	1.00	0.89	0.99	1.00
1	0.97	1.00	1.00	1.00	1.00	1.00	0.97	1.00	1.00
Gaussian EGARCH(1,1) with parameters (-0.001,0.10,0.98,-0.05)									
	$\tau = T/10$			$\tau = T/2$			$\tau = 9T/10$		
ω	T=500	T=1000	T=5000	T=500	T=1000	T=5000	T=500	T=1000	T=5000
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.2	0.11	0.15	0.69	0.51	0.82	1.00	0.10	0.14	0.69
0.4	0.28	0.56	1.00	0.98	1.00	1.00	0.27	0.57	1.00
0.6	0.62	0.95	1.00	1.00	1.00	1.00	0.64	0.95	1.00
0.8	0.91	1.00	1.00	1.00	1.00	1.00	0.92	1.00	1.00
1	0.99	1.00	1.00	1.00	1.00	1.00	0.99	1.00	1.00
Gaussian ARSV(1) with parameters (0.8,0.98,0.02)									
	$\tau = T/10$			$\tau = T/2$			$\tau = 9T/10$		
ω	T=500	T=1000	T=5000	T=500	T=1000	T=5000	T=500	T=1000	T=5000
0	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05	0.05
0.2	0.09	0.15	0.67	0.43	0.78	1.00	0.08	0.14	0.67
0.4	0.24	0.53	1.00	0.95	1.00	1.00	0.21	0.52	1.00
0.6	0.52	0.91	1.00	1.00	1.00	1.00	0.51	0.92	1.00
0.8	0.80	0.99	1.00	1.00	1.00	1.00	0.83	1.00	1.00
1	0.95	1.00	1.00	1.00	1.00	1.00	0.97	1.00	1.00

presence of conditional heteroscedasticity.

Summarizing, for the λ test, the size is larger than the nominal and the power is smaller than in homoscedastic series. However, for the **e** test, both the size and power are similar to the ones obtained in homoscedastic series.

4 Multiple level shifts

The LR-type tests considered before are designed to detect just one level shift at a time. However, in practice, it is possible to encounter real series that contain more than one shift. In this case, Tsay (1988) proposes the following procedure that we denote as **C** (for correct): i) identify the moment of time when the biggest shift occurs, ii) estimate its size and, iii) correct the series for the estimated size. These three steps should be repeated until no further shifts are detected. Then, the joint estimation of all the level shifts detected is recommended by, for example, Chen and Liu (1993). However this procedure can be misleading and inefficient because of the biases of the magnitudes of the shifts estimated in each step. Consider, for example, a series that has two level shifts at times τ_1 and τ_2 respectively. Therefore,

$$z_t = \mu + \omega_1 I(t \geq \tau_1) + \omega_2 I(t \geq \tau_2) + a_t. \tag{9}$$

Without lost of generality, we will consider $\mu = 0$. Suppose that, in the first step, a level shift is detected at time $t = \tau_1$ and its magnitude is estimated. The OLS estimate of ω_1 is given by $\widehat{\omega}_1 = \overline{z}_2 - \overline{z}_1$ where $\overline{z}_1$ and $\overline{z}_2$ are the sample means before and after time τ_1. The expected value of $\widehat{\omega}_1$ is given by:

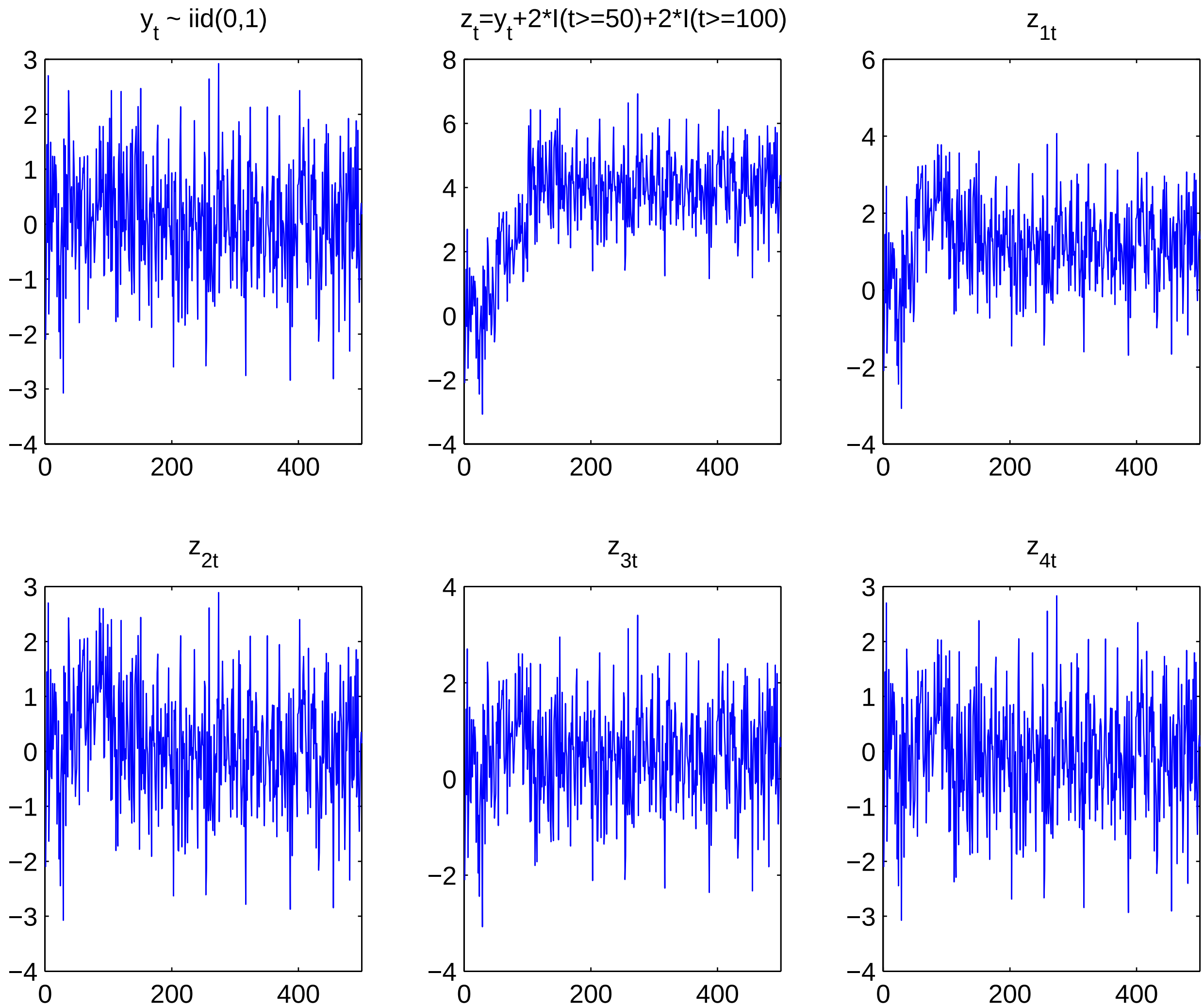
$$E(\widehat{\omega}_1) = \omega_1 + \frac{T - \tau_2 + 1}{T - \tau_1 + 1}\omega_2 \tag{10}$$

Depending on the relationship between the magnitudes and signs of the two level shifts, correcting the original series by $\widehat{\omega}_1$ can generate spurious shifts. To illustrate this problem, Figure 3 plots a zero mean white noise Gaussian series of size $T = 500$, y_t, that has been contaminated with two level shifts of size $w = 2$, in observations $t = 50$ and 100 respectively, obtaining the series $z_t = y_t + 2I(t \geqslant 50) + 2I(t \geqslant 100)$. By applying procedure **C** to the LR test, the original series y_t is only recovered after four corrections. In the first step, $\lambda = |\lambda_{100}| = 23.01$ and, consequently, a level shift is detected at time $t = 100$. The shift size is estimated as $\widehat{\omega}_1 = 2.85$. The corrected series, denoted as z_{1t} is given by $z_{1t} = z_t - 2.85I(t \geqslant 100)$ and has also been plotted in Figure 3. Then, when the LR test is implemented to the series z_{1t}, we obtain $\lambda = |\lambda_{49}| = 7.65$ and a second level shift is detected in the series z_{1t} at time $t = 49$. Its estimated size is $\widehat{\omega}_2 = 1.18$. Once this new level shift is corrected, the new series is given by $z_{2t} = z_{1t} - 1.19I(t \geqslant 49)$. If the λ test is again implemented to the series z_{2t}, the null hypothesis of no level shifts is again rejected with $\lambda = |\lambda_{148}| = 5.20$. Therefore, the third level shift is detected at time $t = 148$ and its estimated size is $\widehat{\omega}_3 = -0.51$. Then, the series $z_{3t} = z_{2t} + 0.51I(t \geqslant 148)$ is obtained and, in this case, $\lambda = |\lambda_{29}| = 3.60$, is again significant and the estimated size is 0.57. Finally, when the λ test is implemented to the series $z_{4t} = z_{3t} - 0.57I(t \geqslant 29)$, the statistic is not significant. If the four level shifts are estimated jointly the result is

$$\widehat{y}_t = \underset{(-0.39)}{-0.08} + \underset{(1.39)}{0.42} I(t \geqslant 29) + \underset{(10.75)}{2.20} I(t \geqslant 49) + \underset{(6.74)}{1.81} I(t \geqslant 100) - \underset{(-2.23)}{0.35} I(t \geqslant 148)$$

The quantities in parenthesis are the t-statistics. Notice that this procedure

Figure 3: Artificial series contaminated with two level shifts and corrected using procedure **C**



can be rather time consuming. Alternatively, it is possible to estimate jointly all the shifts detected up to a particular moment. In our example, if when the second level shift is detected, we estimate jointly the first and second level shift, the estimated model is

$$\widehat{y}_t = \underset{(0.66)}{0.10} + \underset{(12.55)}{1.89}\, I(t \geqslant 49) + \underset{(10.00)}{2.05}\, I(t \geqslant 100)$$

When the λ test is implemented to the residuals of the above model, we obtain $\lambda = 2.17$ which is not significant and, therefore, no spurious level shifts are detected in the corrected series. This problem is the same when the **e** statistic is used.

To illustrate the performance of the **C** procedure, we have simulated 10000 replicates of size $T = 1000$ by model (11) with a_t being a Gaussian white noise process with zero mean and variance one[4]. Each series has been contaminated with two level shifts of the same size $\omega_1 = \omega_2 = 1$, in different positions in the series: $\tau_1 = T/10$, $T/4$, $T/2$, $3T/4$ and $9T/10$ and different distances between shifts $\tau_2 = \tau_1 +\ T/10$ and $\tau_2 = \tau_1 +\ T/4$. Table 9 reports the percentage of rejections of the null hypothesis when the **e** test is implemented to a detect level shift in the contaminated series. The null is rejected in all the simulated series. The second column shows the median through all the Monte Carlo replicates of the period of time when the shift is detected. It can be observed that this time is rather close to one of the actual level shifts. The next two columns of Table 9 report results when the **C** procedure is used after detecting the first level shift. First, we show the percentage

[4]Artificial series have also been generated by conditionally heteroscedastic models with similar results that are not reported here to save space. These results are available from the authors upon request.

Table 9: Comparison of two procedures for testing for multiple level shifts in Gaussian white noise processes using the **e** statistic for $T = 1000$

	Original Series		Procedure **(C)**				Procedure **(D)**			
$\tau_1 - \tau_2$	% reject.	Median t	% reject. Corrected 1	Median t Corrected 1	% reject. Corrected 2	Median t Corrected 2	% reject. Sub-series 1	Median t Sub-series 1	% of reject. Sub-series 2	Median t Sub-series 2
100-200	100	200	86.38	100	91.06	252	100	100	4.44	-
250-350	100	349	99.72	249	99.38	365	100	249	4.24	-
500-600	100	512	94.42	602	84.32	509	30.74	-	88.80	611
750-850	100	750	98.82	849	98.66	717	4.62	-	100	851
800-900	100	799	88.04	899	92.30	747	4.58	-	100	900
100-350	100	349	99.98	101	98.86	371	100	101	4.06	-
250-500	100	497	100	250	100	493	100	250	6.90	-
500-750	100	502	100	749	100	506	7.32	-	100	748

of rejections after each series has been corrected by the estimated shift together with the median time for the second level shift when it is detected. Observe that once the series is corrected by the first level shift detected, the percentage of rejections of the null decreases in some cases even to 84.32. However, if the second shift is detected, the median time of the shift is close to the true time. Finally, the next two columns report the same quantities if a second shift is detected. Notice that, even when the series have not any more level shifts, the percentage of rejections of the null hypothesis is much higher than the nominal size. Therefore, spurious level shifts are detected in a larger number of series. Consider, for example, the case of two level shifts at $\tau_1 = 250$ and $\tau_2 = 350$. In this case, a level shift is detected in median at time 349, corresponding to the second shift. After the series are corrected by the corresponding estimated changes, the null is rejected in 99.72% of the series. The median of the period for the second level shift is 249 corresponding to the first level shift. If the series are again corrected by this second shift, the test rejects the null in 99.38% of the series when there are not more level shifts. The third spurious level shift is detected in median at time $t = 365$. The results for all the other cases considered in Table 9 are similar. In general, we can conclude that the **C** procedure detects correctly the two level shifts but also detects spurious shifts that are not in the original data and apparently occur in moments of time relatively close to the first shift detected. Even if the joint estimation of all the shifts detected is adequate, this procedure is rather inefficient in the sense that it requires quite a lot of steps before the right answer is obtained. The results for the LR test, conditional heteroscedasticity models and for other sample sizes are similar.

Some authors suggest to estimate the second break using all observations but those close to the break previously detected; see, for example, Altissimo and Corradi (2003). However, it is not clear which observations should be taken out when estimating the second break as, in some of the simulated series, we have observed that the spurious breaks can be detected at points which are rather far from the detected shifts.

Given the problems encountered when the **C** procedure is implemented in the presence of two or more level shifts, we consider an alternative procedure that consists on splitting the sample into two subsamples after a level shift has been identified. If, for example, a level shift is detected at time $t = \tau_1$, the series is divided into two subseries: one up to the time $t = \tau_1 - 1$ and the other, from that time on. Then, the test to detect level shifts is implemented in each of the two subseries. The procedure should continue until no further shifts are detected in any of the subseries. The main disadvantage of this procedure, denoted by **D** (for divide), is that, in the successive subdivisions of the original series, the sample sizes of the subseries decrease and consequently, the power of the test also decreases. However, when dealing with financial series, the sample sizes are usually very large and, consequently, in this context, this is a minor problem. Table 9 also reports the results of the Monte Carlo experiments carried out with the same design as before when the procedure **D** is implemented. In this case, after a level shift has been detected, the sample is split into two subsamples. The first and second columns of Table 9 corresponding to procedure **D**, show the percentage of rejections and the median time of the shift in the first subsample and the following two columns are the same quantities for the second subsample. For example, looking at the same case considered above when the shifts occur at times $\tau_1 = 250$ and $\tau_2 = 350$, we observe that once a shift is detected, at time 349, the test detects a second shift in the first subsample in all the simulated series and only in 4.24% of the series in the second subsample, which is close to the nominal size of the test which is 5%. In general, the results reported in Table 8 show that, when two level shifts occur in a time series, the **D** procedure gets quicker to the correct answer than the **C** procedure. It seems that the advantage of the former over the latter procedures will be even more important when more than two shifts occur.

Finally, it is interesting to notice that the **D** procedure seems to work better when the shifts are far apart than when they occur close in time. On the other hand, the procedure works better when the shifts happen in the extremes than when they occur in the middle of the series.

5 Empirical Application

In this section we implement the λ and $\mathbf{e}$ tests to detect level shifts in a series of Spanish Peseta/US Dollar exchange rate returns[5], observed daily from January 2, 1980 to April 18, 2001 with $T = 5371$ observations. The series of returns, y_t, has been plotted in Figure 4 together with a kernel estimate of its density and the correlogram of squared returns. Table 10, that contains some descriptive statistics of y_t, shows that returns exhibit high kurtosis. Furthermore, the statistics proposed by Peña and Rodriguez (2002)

[5]The exchange rates, p_t, have been downloaded from the web page http://pacific.commerce.ubc.ca/xr/ provided by Prof. Werner Antweiler, University of British Columbia, Vancouver, Canada. The series analysed in this paper is the series of returns defined as $y_t = 100(\log(p_t) - \log(p_{t-1}))$.

Figure 4: Exchange rates of Spanish Peseta/ US Dollar observed daily from January, 2, 1980 to April, 18, 2001

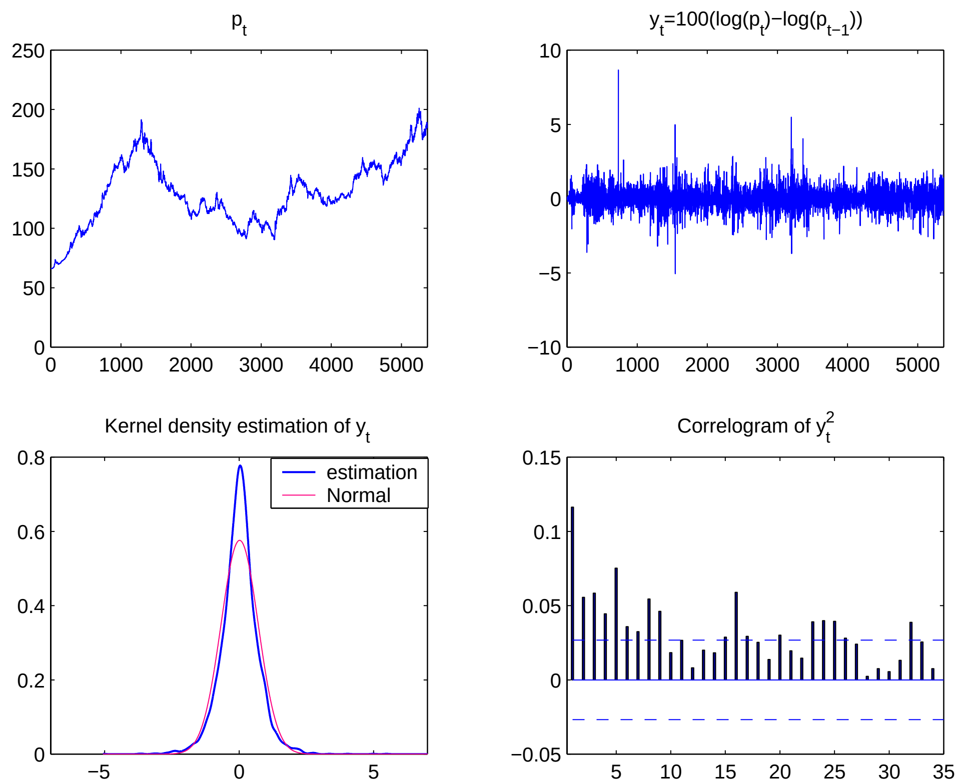


Table 10: Descriptive statistics for the return series y_t and the corresponding sub-series

	y_t	y_{1t}	y_{2t}	y_{3t}
T	5371	1289	1900	2182
	Jan. 2, 1980	Jan. 2, 1980	Feb. 27, 1985	Sept. 3, 1992
	Apr. 18, 2001	Feb. 26, 1980	Sept. 2, 1985	Apr. 18, 2001
Mean	0.02*	0.08*	-0.04*	0.03*
S.D.	0.69	0.67	0.73	0.67
Skewness	0.37*	1.39*	0.06	0.16*
Kurtosis	10.81*	25.18*	6.82*	7.04*
r(1)	-0.02	-0.10*	0.00	-0.02
$D(20)$	11.77	21.43*	16.03	20.52*
Autocorrelation of y_t^2				
$r_2(1)$	0.12*	0.02	0.29*	0.14*
$r_2(2)$	0.06*	0.01	0.09*	0.11*
$r_2(5)$	0.07*	0.02	0.08*	0.20*
$r_2(10)$	0.02	0.00	0.04*	0.04*
$D_2(20)$	129.11*	2.05	179.48*	167.74*

T: Sample size.
$r(\tau)$: Order τ autocorrelation of y_t.
$r_2(\tau)$: Order τ autocorrelation of y_t^2.
*Statistically significant at 95% of confidence.

to test for uncorrelatedness of y_t and y_t^2, $D(k)$ and $D_2(k)$ respectively, shows that although the series y_t is uncorrelated, the autocorrelations of squared observations are significantly different from zero. These are properties that usually characterize conditional heteroscedasticity.

First, we test for a level shift in the series of returns using the LR test. Figure 5 plots the values of the λ_m statistic, $m = 2, ..., 5371$ wich has a maximum $\lambda = |\lambda_{1289}| = 3.75$ which is larger than 3.43, the 5% critical value for $T = 1500$ in Table 1. Therefore, a level shift is detected at time $t = 1289$ that corresponds to February, 26, 1985. Notice that in 1985 the G5 decided, in Washington D.C., devaluate the Dollar with the objective of improve exportations. Once the shift is detected, we estimate its magnitude by OLS with the following results:

$$\widehat{y}_t = \underset{(4.26)}{0.08} - \underset{(-3.72)}{0.08} \, I(t \geq 1289) \tag{11}$$

Then, the series is corrected and the test is applied again to look for another shift. Figure 5, also shows the new values of the λ_m statistic applied to the corrected series. In this case, no more changes are detected.

We also compute the λ_m statistic in each of the two subsamples obtained splitting the sample before and after February, 26, 1985. In this case, a new shift is detected at time $t = 1290$, i.e. in the first observation of the second subseries. If the test is applied to the subseries y_{2t} for $t = 1291, \ldots, 5371$ no more shifts are detected.

Alternatively, we test for level shifts in the returns series implementing the $\mathbf{e}$ statistic. Figure 6 plots the values of e_m that reaches a maximum at time $t = 1289, \mathbf{e} = |e_{1289}| = 1.60$ which is larger than 1.4, the 5% asymptotic

Figure 5: Values of the λ statistic for the Spanish Peseta/US Dollar returns and the corrected series

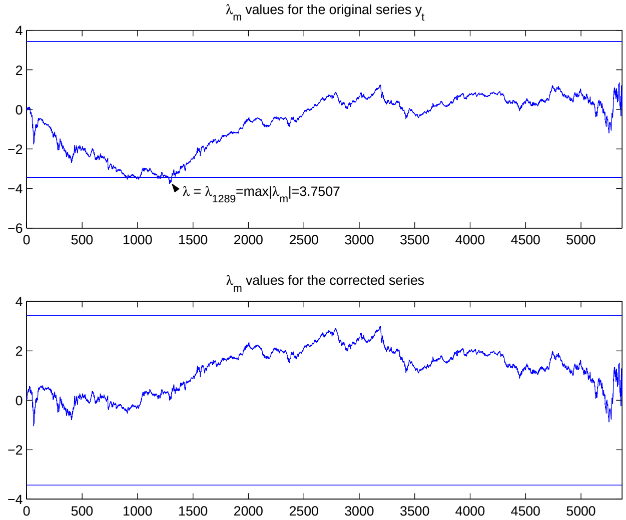
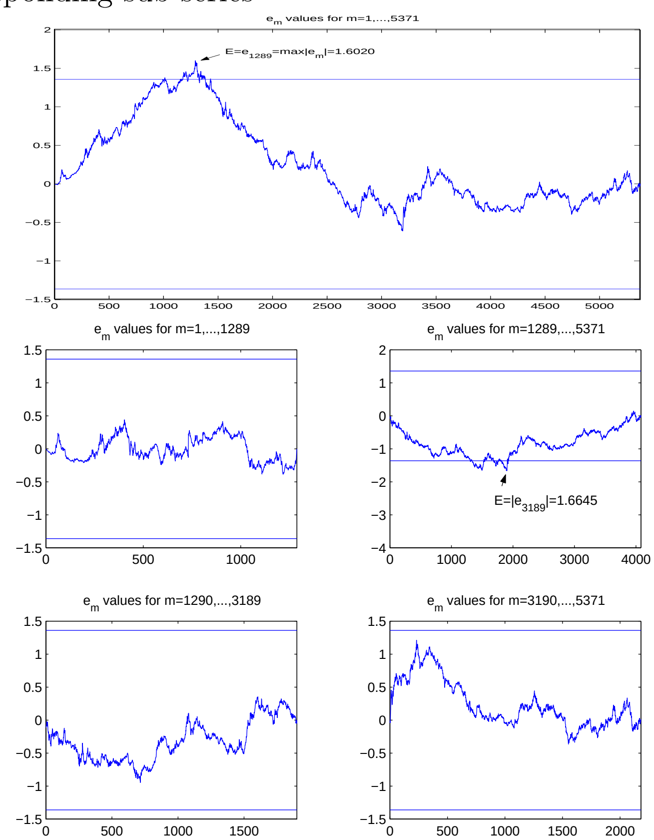


Figure 6: Values of the **e** statistic for the Spanish Peseta/US Dollar returns and the corresponding sub-series



critical value. Therefore, the null hypothesis is rejected, indicating that there is a level shift at February, 26, 1985, in agreement with the LR test. Then, the series is split into two subseries y_{1t} and y_{2t}, the first one from $t = 1$ up to $t = 1289$ and the second from $t = 1290$ up to $t = 5371$. Then, the procedure is applied again to each of the subseries. As we can see in Figure 6, the values of e_m do not cross the critical value in the first subsample, while, in the second subsample, the maximum is $\mathbf{e} = |e_{3189}| = 1.66$ which is larger than the critical value and, consequently, a new level shift is detected at time $t = 3189$ corresponding to September, 2, 1992. Notice that, in September 1992 the Peseta was devaluated several times. If we apply again the procedure to the two new subseries: the first one from $t = 1290$ up to $t = 3189$ and the second from $t = 3190$ up to $t = 5371$, no more level shifts are detected. As Table 10 shows, the means of the three subseries are different.

If the $\mathbf{e}$ test is implemented after correcting the series, it detects another shift at time $t = 3189$. Therefore, for this particular example, the statistic $\mathbf{e}$ detects two level shifts and λ detects just one shift independently of the procedure used. This result agrees with the Monte Carlo results that show the lack of power of the λ test to detect level shifts in the presence of conditional heteroscedasticity.

6 Conclusions

In this paper we have studied the properties of two variants of the LR statistic for detecting a level shift in uncorrelated conditionally heteroscedastic time series. We show that while the standard LR test, λ, suffers from important size distortions, the **e** test is robust, at least when the conditional heteroscedasticity is generated by some of the most popular models in the literature. Furthermore, the **e** test does not lose power while the λ test may have important decreases in power and, therefore may have problems to detect shifts actually present in the series.

We have also compared two procedures for detecting multiple level shifts. When a level shift is detected, the first one corrects the series by the estimated size, whereas the second divides the series at the time detected. For the large sample sizes usually encountered when analyzing financial time series, the second procedure seems to get quicker to the right answer.

Finally, both tests are applied to a daily series of returns of the Spanish Peseta/US Dollar exchange rates. In this particular series, the λ test only detects one shift while the **e** test finds two shifts that are justified by the characteristics of the series analyzed.

References

Altissimo, F. and V. Corradi (2003), Strong rules for detecting the number of breaks in a time series, *Journal of Econometrics,* 117, 207-244.

Andrews, D.W.K. (1993). Tests for Parameter Instability and Structural Change with Unknown Change Point. *Econometrica* , 61, 821-856.

Bai, J. (1994). Least Squares Estimation of a Shift in Linear Processes. *Journal of Time Series Analysis,* 15, 453-472.

Balke, N.S. (1993). Detecting level Shifts in Time Series. *Journal of Business and Economic Statistics*, 11, 81-92.

Bollerslev, T., R.F. Engle and D.B. Nelson (1994) ARCH Models. *The Handbook of Econometrics*, 4, 2959-3038.

Carnero, M.A., D. Peña and E. Ruiz (2001). Outliers and Conditional Autoregressive Heteroscedasticity, *Estadística*, 53, 143-213.

Carnero, M.A., D. Peña and E. Ruiz (2003). Why is GARCH more persistent and conditionally leptokurtic than Stochastic Volatility?. manuscript, Universidad Carlos III de Madrid.

Ghysels, E., A. Harvey and E. Renault (1996). Stochastic Volatility. *Handbook of Statistics*, 14.

Hawkins, D.M. (1977). Testing a Sequence of Observations for a Shift in Location. *Journal of the American Statistical Association,* 72, 180-186.

Inclan, C. and G.C. Tiao (1994). Use of Cumulative Sums of Squares for retrospective Detection of Changes of Variance. *Journal of the American*

Statistical Association, 89, 913-923.

Krämer, W., W. Ploberger and R. Alt (1988). Testing for Structural Change in Dynamic Models. *Econometrica*, 56, 1355-1369.

Peña, D. and Rodriguez, J. (2002). A Powerful Portmanteau Test of Lack of Fit in Time Series, *Journal of the American Statistical Association*, 97.

Shephard, N.G. (1996). Statistical Aspects of ARCH and Stochastic Volatility. In *Time Series Models in Econometrics, Finance and Other Fields,* D.R. Cox, D.V. Hinkley and O.E. Bardorff-Nielsen (eds.). Chapman & Hall, London.

Tsay, R.S. (1988). Outliers, Level Shifts and Variance Changes in Time Series. *Journal of Forecasting*, 7, 1-20.

Worsley, K.J. (1986). Confidence Regions and Tests for a Change-Point in a Sequence of Exponential Family Random Variables. *Biometrika,* 73, 91-104.