SPECIFICATION TESTS FOR THE ERROR DISTRIBUTION IN GARCH MODELS

B. Klar[a], F. Lindner[a], S.G. Meintanis[b,*]

[a] *Institut für Stochastik, Karlsruhe Institute of Technology (KIT), Kaiserstraße 89, 76133 Karlsruhe, Germany*
[b] *Department of Economics, National and Kapodistrian University of Athens, 8 Pesmazoglou Street, 105 59 Athens, Greece*

Abstract

Goodness–of–fit and symmetry tests are proposed for the innovation distribution in generalized autoregressive conditionally heteroscedastic models. The tests utilize an integrated distance involving the empirical characteristic function (or the empirical Laplace transform) computed from properly standardized observations. A bootstrap version of the tests serves the purpose of studying the small sample behaviour of the proclaimed procedures in comparison with more classical approaches. Finally, all tests are applied to some financial data sets.

Keywords: GARCH model, Goodness–of–fit test, Symmetry test, Empirical characteristic function, Bootstrap test.

1. Introduction

Suppose that a process is driven by a generalized autoregressive conditionally heteroscedastic (GARCH) model of specific order. This class of models was introduced by Bollerslev (1986), and despite the fact that certain properties of GARCH processes make no use of the particular form of the innovation distribution (see, for example, the stationarity conditions mentioned later in this section), GARCH models with Gaussian innovations have been considered. However, on the basis of numerous real–data applications accumulated over the years, the normality assumption has become questionable, not only for the marginal, but also for the conditional distribution of certain

*Corresponding author. *Tel./Fax.* +302103689814
Email address: simosmei@econ.uoa.gr (S.G. Meintanis)

financial quantities. Specifically, it often appears that such quantities are skewed and contain a persistent amount of leptokurtosis; refer, for instance, to Bollerslev (1987), Nelson (1991), Granger and Ding (1995), Curto et al. (2009), Mittnik and Paolella (2003), and Mittnik et al. (1998).

Although maximum–likelihood type estimators for GARCH models are consistent and asymptotically normal under mild conditions, regardless of the error distribution, the impact of an incorrect Gaussian assumption has been investigated by several researchers. Engle and González-Rivera (1991) show that there is a loss of efficiency as high as 84% of the so–called Quasi MLE (QMLE) of the parameters under non–normal innovations. González-Rivera and Drost (1999) theoretically show that the efficiency of the QMLE and a certain semiparametric estimator, relative to the MLE, depends on the kurtosis as well as the skewness of the conditional error–density. In fact, and unlike the case of symmetry which allows for equal efficiency of the three estimators, the presence of skewness rules out the possibility of a fully efficient QML estimator. On the other hand, a QMLE associated with an incorrect non–Gaussian specification may also imply inconsistency. Specifically, Newey and Steigerwald (1997) show that in certain conditional autoregressive models nesting the GARCH, inconsistency may result not from the presence of non–Gaussian errors alone, but if either the assumed or the true error density is asymmetric. For simulation results on the behaviour of the QML estimator under misspecified GARCH(1,1) models see Bellini and Bottolo (2009). Further evidence from Huang et al. (2008) suggests that least absolute deviations estimators may be preferred to maximum likelihood estimation under Laplace and certain Student's–t innovation distributions. There is also considerable evidence on the impact of correct specification on other aspects of modelling, such as predictions; see, for instance, Hansen and Lunde (2005), Forsberg (2002) and Bellini and Bottolo (2009).

Despite the aforementioned importance of the innovation distribution, literature contains few references to corresponding specification tests. Kulperger and Yu (2005), for example, modify the well-known test of Jarque and Bera (1987) (based on skewness and kurtosis) with respect to the case of GARCH specification in order to test normality of the innovation distribution. Horváth and Zitikis (2006) propose a general goodness–

of–fit test based on a nonparametric estimator of the innovation density and Horváth et al. (2004) suggest certain modifications of classical statistics based on the distribution function of squared innovations, while Koul and Ling (2006) propose a weighted version of the Kolmogorov–Smirnov test which has a model–free asymptotic null distribution.

In this paper we propose goodness–of–fit tests for the innovation distribution based on the 'Fourier approach'. This approach utilizes the characteristic function in order to test the corresponding null hypothesis. Specifically, we consider test statistics for GARCH(p,q) models which are based on the empirical characteristic function (ECF)

$$\hat{\varphi}_T(u) = \frac{1}{T-m} \sum_{t=m+1}^{T} e^{iu\hat{\varepsilon}_t}, \quad -\infty < u < \infty, \ \hat{\varepsilon}_t = y_t/\hat{c}_t, \tag{1.1}$$

where $m = \max(p,q)$. The corresponding observations $y_1, ..., y_T$ are from the GARCH model of fixed order (p,q)

$$y_t = c_t\varepsilon_t, \quad c_t^2 = \beta_0 + \sum_{j=1}^{q} \beta_j y_{t-j}^2 + \sum_{j=1}^{p} \gamma_j c_{t-j}^2,$$

$$\beta_0 > 0, \quad \beta_j \geq 0 \ (1 \leq j \leq q), \quad \gamma_j \geq 0 \ (1 \leq j \leq p),$$

incorporating the i.i.d. innovations $\varepsilon_t, \ t = 1, ..., T$, with the empirical scales $\hat{c}_t$ being computed from an estimate of the unknown parameter–vector $\vartheta = (\beta_0, ..., \beta_q, \gamma_1, ...\gamma_p)$. The GARCH$(p,q)$ model is covariance stationary if and only if $\sum_{j=1}^{q} \beta_j + \sum_{j=1}^{p} \gamma_j < 1$. Strict stationarity conditions have also been derived by Bougerol and Picard (1992a,b). Considering the GARCH(1,1) model, for example, the necessary and sufficient condition reads as $\mathbf{E}[\log(\gamma_1 + \beta_1\varepsilon^2)] < 0$.

The remaining paper unfolds as follows. In Section 2, we introduce the tests and discuss some aspects of the test statistics, as well as the important issue of estimation of parameters. Bootstrap versions of the tests are introduced in Section 3 and their behaviour is studied by means of Monte Carlo. In Section 4, we apply the methods to some real data, and finally our findings are summarized in Section 5.

2. Test statistics and estimation

(i) Test statistics. The test statistics based on the ECF are analogous to the corresponding goodness–of–fit tests for i.i.d. observations. In particular, $F(\cdot)$ and $\varphi(\cdot)$

denote the distribution function and the characteristic function of the innovations, respectively. Despite the fact that only tests for specific distributions are considered in the following, we formulate the null hypothesis as $H_0 : F \in \mathcal{F}$, where $\mathcal{F}$ denotes a parametric family of distributions. Thus, the hypothesized distribution is allowed to depend on an unknown parameter, which ought to be estimated from the data; we shall come back to this issue in the conclusion. The test statistic then takes the form

$$\Phi_{T,w} = (T - m) \int_{-\infty}^{\infty} |\hat{\varphi}_T(u) - \varphi(u)|^2 w(u) du, \tag{2.1}$$

with $w(u)$ denoting an appropriate weight function introduced in order to taper the persistent periodic behaviour of $\hat{\varphi}_T(u)$. Note that in the context of i.i.d. observations, the limit behaviour of the ECF was systematically studied in a series of papers by Feuerverger and Mureika (1977), Csörgő (1981), Marcus (1981), and Csörgő and Totik (1983). The amiable limit properties of the ECF have furthered the use of L2–test statistics in the style of eqn. (2.1). In turn, the theoretical properties of these statistics have been investigated by Henze and Wagner (1997), Meintanis (2004), Epps (2005) and Matsui and Takemura (2007) for goodness-of–fit tests with some particular cases of parametric distributions, and by Meintanis and Swanepoel (2007) under a general formulation in a Hilbert space setting, including assumptions on the weight function $w(u)$ and the estimators of parameters involved. However, concerning the actual performance of the tests, the limit null distribution derived is extremely difficult to employ, and we do not pursue this issue any further at this point. On the other hand, these earlier studies have shown that the ECF statistics can be more powerful and/or more convenient to use than the more standard procedures.

To gain some insight into the method proposed, suppose that $w(-u) = w(u)$ and write $C(u)$ (resp. $S(u)$) for the real part (resp. imaginary part) of the characteristic function of the innovations $\varphi(u)$. The test statistic then may be written as

$$\Phi_{T,w} = (T - m) \int_{0}^{\infty} g(u) w(u) du, \tag{2.2}$$

where $g(u) = [(C_T(u) - C(u)) + (S_T(u) - S(u))]^2 + [(C_T(u) - C(u)) - (S_T(u) - S(u))]^2$ involves the real part $C_T(u) = (T - m)^{-1} \sum_{t=m+1}^{T} \cos(u\hat{\varepsilon}_t)$ and the imaginary part $S_T(u) = (T - m)^{-1} \sum_{t=m+1}^{T} \sin(u\hat{\varepsilon}_t)$ of the ECF. Assuming the existence of enough moments ($\mathbf{E}(\varepsilon^r) < \infty$) for the innovation distribution and employing Taylor expansions of $\cos(\cdot)$ and $\sin(\cdot)$ yields after some algebra

$$\begin{aligned}\Phi_{T,w} &= (T-m)\Bigg[\kappa_2 2M_1^2 + \kappa_4\left(\frac{1}{2}M_2^2 - \frac{2}{3}M_1M_3\right) + \\ &\quad + \kappa_6\left(\frac{1}{30}M_1M_5 - \frac{1}{12}M_2M_4 + \frac{1}{18}M_3^2\right) + \ldots\Bigg], \end{aligned} \tag{2.3}$$

where $\kappa_r := \int_0^\infty u^r w(u)du$, $r = 2,4,6,\ldots$, and $M_r := (T-m)^{-1}\sum_{t=m+1}^T \hat{\varepsilon}_t^r - \mathbf{E}(\varepsilon^r)$. This shows that in (2.3) moment–matching takes places – the sample moments based on $\hat{\varepsilon}_t$ on the one hand, the theoretical moments of the hypothesized distribution on the other. A typical choice for $w(\cdot)$ is an exponentially decaying function, such as $w(u) = e^{-a|u|^b}$, $a, b > 0$, which can be easily seen from (2.3) to yield the limiting values $\lim_{a\to\infty} a^3\Phi_{T,w} = 4(T-m)M_1^2$, and $\lim_{a\to\infty} a^{3/2}\Phi_{T,w} = \sqrt{\pi/2}\,(T-m)M_1^2$, for $b = 1$ and $b = 2$, respectively.

(ii) Estimation of GARCH models. In the Gaussian GARCH(p,q) model, estimation of ϑ may be accomplished by maximizing the conditional likelihood,

$$L_T(\vartheta) = -\frac{1}{2}\sum_{t=m+1}^{T}\left[\log(c_t^2) + \frac{y_t^2}{c_t^2}\right]. \tag{2.4}$$

Berkes et al. (2003) and Francq and Zakoïan (2004) provide recursive representations of the conditional variance c_t^2. Given initial values, these representations may be used in the estimation procedure. Under positivity constraints for the GARCH coefficients, Berkes et al. (2003) proved that for the Gaussian GARCH as well as for an arbitrary GARCH(p,q) model with $\mathbf{E}(|\varepsilon^2|^{2+\delta}) < \infty$, for some $\delta > 0$, the resulting quasi maximum likelihood estimator is consistent and asymptotically normal. Francq and Zakoïan (2004) proved the asymptotic normality result under the slightly milder condition $\mathbf{E}(\varepsilon^4) < \infty$, whereas Berkes and Horváth (2004) – by replacing the normal function in the QMLE by the Laplace density – obtain consistency and asymptotic normality only under $\mathbf{E}(\varepsilon^2) < \infty$. For asymptotic properties of the fully parametric MLE (even with infinite–variance innovations) and the loss of efficiency due to the use of the QMLE compared to the MLE the reader is referred to Francq and Zakoïan (2006).

(iii) Statistic for testing symmetry. For testing the hypothesis that the error distribution is symmetric about the origin, we use the statistic

$$\Sigma_{T,w} = \int_{-\infty}^{\infty} \left((T-m)^{-1/2} \sum_{t=m+1}^{T} \sin(u\hat{\varepsilon}_t) \right)^2 w(u)du, \tag{2.5}$$

based on the imaginary part of the ECF. In the following, the weight function $w(u)$ is chosen to be the density of the standard normal distribution. Since the distribution of a random variable X is symmetric about 0 if, and only if, the imaginary part of the characteristic function of X vanishes, i.e., if $\mathbf{E}[\sin(uX)] = 0$ for each $u \in \mathbb{R}$, this statistic seems very natural. It was first considered by Feuerverger and Mureika (1977) in the i.i.d. case. For adaptation of analogous test statistics in the context of time series the reader is referred to Ngatchou-Wandji (2009).

One could again choose other suitable weight functions in the definition of the statistic $\Sigma_{T,w}$ in (2.5); the above choice leads to the simple explicit form

$$\Sigma_{T,w} = \frac{1}{2(T-m)} \sum_{s,t=m+1}^{T} \left[\exp\left(-(\hat{\varepsilon}_s - \hat{\varepsilon}_t)^2/2\right) - \exp\left(-(\hat{\varepsilon}_s + \hat{\varepsilon}_t)^2/2\right) \right].$$

3. Simulations

3.1. Goodness of fit tests for the error distribution

In this section the finite–sample behaviour of several goodness–of–fit tests is studied by Monte Carlo. In particular, the following bootstrap procedure is employed in order to compute the critical point of the test:

(i) On the basis of of $y_1, ..., y_T$, compute the estimator $\hat{\vartheta}_T$ of the parameter–vector ϑ and $\hat{c}_t := c_t(\hat{\vartheta}_T)$.

(ii) Compute the test statistic $\Phi_{T,w} := \Phi_{T,w}(\widehat{\mathbf{E}}_T)$ where $\widehat{\mathbf{E}}_T$ denotes a vector with elements $\hat{\varepsilon}_t, \ t = m+1, ..., T$.

(iii) Define the bootstrap observations

$$y_t^* = \hat{c}_t^* \varepsilon_t^*, \quad (\hat{c}_t^*)^2 = \hat{\beta}_0 + \sum_{j=1}^{q} \hat{\beta}_j (y_{t-j}^*)^2 + \sum_{j=1}^{p} \hat{\gamma}_j (\hat{c}_{t-j}^*)^2,$$

where $\varepsilon_t^*, \ t = 1, ..., T$, are i.i.d. observations from the hypothesized innovation distribution.

(iv) Based on $(y_1^*, \ldots, y_T^*)$, compute the estimates ϑ_T^* and $c_t^* := \hat{c}_t^*(\vartheta_T^*)$.
(v) Compute the test statistic $\Phi_{T,w}^* := \Phi_{T,w}(\mathbf{E}_T^*)$, where $\mathbf{E}_T^*$ denotes a vector with elements $y_t^*/c_t^*,\ t = m+1, ..., T$.

When steps (iii)–(v) are repeated a number of times, say B, the sampling distribution of $\Phi_{T,w}$ is reproduced, and on the basis of this bootstrap distribution we decide whether the observed value of the test statistic is significant. These bootstrap schemes were employed by Horváth et al. (2004).

In the following, we solely consider the hypothesis of normal innovations $\varepsilon_t \sim N(0,1)$, additional null hypotheses will be examined in Section 4. In the case of normality, any empirical distribution function (EDF) test statistic can be based on the usual empirical process of residuals

$$\tilde{\alpha}_T(x) = \frac{1}{\sqrt{T-m}} \sum_{t=m+1}^{T} \left(I[\hat{\varepsilon}_t \leq x] - N(x) \right), \quad -\infty < x < \infty, \tag{3.1}$$

where N is the standard normal cumulative distribution function, or on the empirical process of squared residuals

$$\hat{\alpha}_T(x) = \frac{1}{\sqrt{T-m}} \sum_{t=m+1}^{T} \left(I[\hat{\varepsilon}_t^2 \leq x] - (2N(\sqrt{x}) - 1) \right), \quad x > 0. \tag{3.2}$$

Likewise, ECF test statistics can make use of both types of residuals.

Horváth et al. (2004) considered three test statistics, all based on the empirical process $\hat{\alpha}_T$ of squared residuals:

- A Cramér-von Mises type statistic (here denoted by HKT-CM-sr)

  $$CM_{T,1} = \int |\hat{\alpha}_T(x)|^2 \, dx.$$

  (Here, the suffix -sr stands for squared residuals).

- A further normalized variant of the Cramér-von Mises statistic (here denoted by HKT-NCM-sr)

  $$CM_{T,2} = \int |\hat{\alpha}_T(x)|^2 \, \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \, dx.$$

	H_0			H_1			H_2		
	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
EP	0.094	0.047	0.008	1.000	1.000	0.998	0.103	0.054	0.008
CVM	0.099	0.047	0.008	0.997	0.979	0.820	0.093	0.050	0.010
HKT-CM	0.093	0.047	0.008	0.987	0.935	0.703	0.094	0.055	0.012
HKT-NCM	0.098	0.045	0.008	0.938	0.874	0.639	0.094	0.055	0.013
KS	0.098	0.048	0.008	0.999	0.986	0.843	0.105	0.048	0.010
HKT-KS	0.096	0.047	0.008	0.978	0.949	0.829	0.101	0.054	0.008
AD	0.102	0.048	0.009	1.000	1.000	0.973	0.104	0.051	0.011
W	0.096	0.049	0.011	1.000	1.000	0.999	0.132	0.062	0.012
LT-sr	0.093	0.050	0.012	0.772	0.606	0.329	0.190	0.118	0.036
CVM-sr	0.097	0.054	0.013	0.808	0.652	0.336	0.176	0.102	0.031
HKT-CM-sr	0.092	0.044	0.010	0.992	0.980	0.905	0.211	0.133	0.038
HKT-NCM-sr	0.010	0.048	0.011	0.823	0.688	0.416	0.199	0.115	0.030
KS-sr	0.010	0.048	0.010	0.826	0.687	0.398	0.165	0.091	0.023
HKT-KS-sr	0.101	0.048	0.010	0.846	0.714	0.425	0.187	0.107	0.030
AD-sr	0.096	0.053	0.013	0.906	0.796	0.466	0.172	0.100	0.031
W-sr	0.111	0.052	0.013	0.865	0.772	0.497	0.146	0.076	0.017

Table 1: Empirical frequencies of rejection based on 2500 replications, $T = 400$, $B = 199$

- A one-sided Kolmogorov-Smirnov statistic (here denoted by HKT-KS-sr)

$$KS_T = \max_{m+1 \leq t \leq T} \left| \hat{\alpha}_T(\hat{\varepsilon}_t^2) \right| .$$

We further included the usual Kolmogorov-Smirnov (KS-sr), Cramér-von Mises (CM-sr), Anderson-Darling (AD-sr) and Watson (W-sr) statistics (see D'Agostino and Stephens (1986)) in our simulation study, all of those, again, based on the squared residuals. For the last three statistics, very simple explicit formulae exist, whereas HKT-CM-sr and HKT-NCM-sr have to be computed by numerical integration.

The same statistics, but applied to the empirical process of non-squared residuals $\tilde{\alpha}_n$, are denoted by the same corresponding abbreviations, but without the suffix -sr.

The Anderson-Darling test is the EDF test for normality recommended by D'Agostino and Stephens (1986).

Finally, we included the Epps-Pulley test statistic (EP) based on $\tilde{\alpha}_n$ as a representative of ECF tests for normality Epps and Pulley (1983). This statistic may be computed from (2.1) by using the ECF in (1.1) and $\varphi(u) = e^{-(1/2)u^2}$, and by setting the weight function $w(u)$ equal to the standard normal density. (Refer to the weight function $w(u) = e^{-a|u|^b}$ at the end of Section 3 and let $b = 2$ and $a = 1/2$). The above choice leads to the simple explicit form

$$\Phi_{T,w} = \frac{1}{(T-m)} \sum_{s,t=m+1}^{T} \exp\left(-(\hat{\varepsilon}_s - \hat{\varepsilon}_t)^2/2\right) - \sqrt{2} \sum_{t=m+1}^{T} \exp\left(-\hat{\varepsilon}_t^2/4\right) + \frac{(T-m)}{\sqrt{3}}.$$

One could define a similar test statistic based on $\hat{\alpha}_n$, but this approach does not lead to an explicit expression. Hence, we decided to base a test on the empirical Laplace transform in the following way:

For $u > 0$, the Laplace transform $L(u) = \int_0^\infty e^{-uy} dF(y) = (1+2u)^{-1/2}$ of a chi-squared random variable with one degree of freedom satisfies

$$(1+2u)L'(u) + L(u) = 0.$$

Now, let $x_1, \ldots, x_n$ be independent observations supposedly following a χ_1^2-distribution, and let $L_n(u) = n^{-1} \sum_{j=1}^n e^{-ux_j}$ denote the empirical Laplace transform of $x_1, \ldots, x_n$. Hence, a suitable test statistic is given by

$$\begin{aligned} \Psi_{n,w} &= n \int_0^\infty \left((1+2u)L_n'(u) + L_n(u)\right)^2 w(u)du \\ &= \frac{1}{n} \sum_{j,k=1}^{n} \int_0^\infty \left\{(1+2u)^2 x_j x_k - (1+2u)(x_j + x_k) + 1\right\} e^{-u(x_j+x_k)} w(u)du, \end{aligned}$$

with some positive weight function w. Choosing $w(u) = e^{-au}$ with $a \geq 0$, we obtain

$$\Psi_{n,a} = \frac{1}{n} \sum_{j,k=1}^{n} \left\{ \frac{(x_j + x_k + a + 2)^2 + 4}{(x_j + x_k + a)^3} x_j x_k - \frac{(x_j + x_k + a + 2)}{(x_j + x_k + a)^2} (x_j + x_k) + \frac{1}{x_j + x_k + a} \right\}.$$

The test corresponding to $\Psi_{n,1}$ where n is replaced by $(T-m)$, and $x_1, \ldots, x_n$ are replaced by the squared residuals $\hat{\varepsilon}_{m+1}^2, \ldots, \hat{\varepsilon}_T^2$, is denoted by LT-sr in our simulations.

	H_3			H_4			H_6		
	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
EP	0.138	0.063	0.010	0.612	0.395	0.091	0.998	0.988	0.835
CM	0.120	0.058	0.010	0.322	0.140	0.019	0.989	0.923	0.534
HKT-CM	0.124	0.075	0.018	0.327	0.192	0.052	0.852	0.741	0.463
HKT-NCM	0.125	0.071	0.017	0.282	0.161	0.044	0.818	0.701	0.411
KS	0.138	0.068	0.013	0.349	0.193	0.039	0.972	0.912	0.599
HKT-KS	0.130	0.069	0.013	0.310	0.166	0.040	0.852	0.731	0.438
AD	0.119	0.059	0.012	0.418	0.209	0.034	0.993	0.951	0.618
W	0.202	0.109	0.024	0.755	0.612	0.277	1.000	1.000	0.996
LT-sr	0.329	0.218	0.080	0.898	0.823	0.530	1.000	1.000	1.000
CM-sr	0.297	0.204	0.075	0.880	0.793	0.493	1.000	1.000	1.000
HKT-CM-sr	0.314	0.218	0.082	0.905	0.835	0.562	1.000	1.000	0.999
HKT-NCM-sr	0.308	0.216	0.081	0.892	0.818	0.538	1.000	1.000	1.000
KS-sr	0.266	0.175	0.059	0.830	0.729	0.412	1.000	1.000	1.000
HKT-KS-sr	0.284	0.194	0.067	0.848	0.747	0.430	1.000	1.000	1.000
AD-sr	0.294	0.196	0.073	0.879	0.792	0.488	1.000	1.000	1.000
W-sr	0.218	0.130	0.037	0.722	0.590	0.284	1.000	1.000	0.998

Table 2: Empirical frequencies of rejection based on 2500 replications, $T = 400$, $B = 199$

In Tables 1-3, the upper (lower) parts show the simulation results for the test statistics based on non-squared (squared) residuals, considering the following hypothetical and alternative data generating processes (see also Horváth et al. (2004)). The data follow an ARCH(2) model

$$y_t = c_t \varepsilon_t, \quad c_t^2 = 0.1 + 0.2 y_{t-1}^2 + 0.1 y_{t-2}^2.$$

With the exception of alternative H_3, the innovations are independent. Under the hypothesis H_0, all innovations have a standard normal distribution. Under the alternatives, innovations follow other distributional models, and/or there is a change point in the series. More specifically, the innovations for alternatives H_1 and H_2 are given by

$$\varepsilon_t = \frac{\eta_t - 5}{\sqrt{10}},\ \eta_t \sim \chi_5^2,$$

and

$$\varepsilon_t \sim N(0,1),\ t \leq T/2, \quad \varepsilon_t \sim N(0,2),\ T > T/2,$$

respectively. The model used in H_3 is an ARCH(2) process with innovations following a linear ARCH(1) process as introduced by Giraitis et al. (2000); this is DGP 3 in Horváth et al. (2004). The innovations in the remaining alternatives are defined by

$$\varepsilon_t = \sqrt{\frac{6}{4}} u_t,\ u_t \sim t(6), \qquad (H_4)$$

$$\varepsilon_t \sim N(0,1),\ t \leq T/2, \quad \varepsilon_t = \frac{\eta_t - 5}{\sqrt{10}},\ \eta_t \sim \chi_5^2,\ t > T/2, \qquad (H_5)$$

$$\varepsilon_t \sim \mathit{Laplace}\left(1/\sqrt{2}\right), \qquad (H_6)$$

$$\varepsilon_t \sim N(0,1),\ t \leq T/2, \quad \varepsilon_t \sim \mathit{Laplace}\left(1/\sqrt{2}\right),\ t > T/2 \qquad (H_7)$$

and

$$\varepsilon_t \sim N(0,1),\ t \leq T/2, \quad \varepsilon_t \sim t(7),\ t > T/2 \qquad (H_8).$$

Tables 1-3 show empirical levels and powers of the different tests based on 2500 replications; in each case, $B = 199$ bootstrap samples have been drawn. Columns 2-4 of Table 1 demonstrate that all bootstrap tests maintain the theoretical level closely – at least for the sample size $T = 400$, which we used in all simulations. For most alternatives, a clear-cut differentiation is possible between the tests based on squared and non-squared residuals. Each group is preferable for certain alternatives.

Within the group of tests based on non-squared residuals, the Epps-Pulley and the Watson test are most powerful. In the other group, the Cramér-von Mises type tests HKT-CM-sr and HKT-NCM-sr are very powerful; the test based on the empirical Laplace transform and the Anderson-Darling test have comparably high power for most alternatives.

The second simulation utilizes a GARCH(1,1) process

$$y_t = c_t \varepsilon_t, \quad c_t^2 = 0.1 + 0.3 y_{t-1}^2 + 0.3 c_{t-1}^2.$$

	H_7			H_8			H_5		
	0.1	0.05	0.01	0.1	0.05	0.01	0.1	0.05	0.01
EP	0.566	0.347	0.080	0.226	0.121	0.022	0.851	0.717	0.410
CVM	0.373	0.183	0.026	0.129	0.062	0.014	0.552	0.390	0.163
HKT-CM	0.344	0.230	0.070	0.150	0.084	0.022	0.447	0.288	0.100
HKT-NCM	0.328	0.218	0.066	0.134	0.074	0.019	0.421	0.286	0.107
KS	0.451	0.268	0.076	0.149	0.079	0.017	0.565	0.424	0.194
HKT-KS	0.367	0.228	0.068	0.148	0.079	0.017	0.632	0.502	0.248
AD	0.431	0.224	0.038	0.162	0.074	0.016	0.741	0.522	0.177
W	0.847	0.747	0.461	0.308	0.197	0.069	0.899	0.829	0.609
LT-sr	0.948	0.911	0.788	0.487	0.381	0.206	0.330	0.206	0.077
CM-sr	0.940	0.902	0.766	0.447	0.338	0.169	0.333	0.210	0.068
HKT-CM-sr	0.916	0.867	0.711	0.546	0.404	0.255	0.663	0.545	0.293
HKT-NCM-sr	0.930	0.886	0.728	0.481	0.345	0.199	0.376	0.258	0.093
KS-sr	0.915	0.849	0.652	0.393	0.283	0.120	0.304	0.190	0.050
HKT-KS-sr	0.927	0.870	0.681	0.431	0.308	0.130	0.319	0.200	0.057
AD-sr	0.940	0.906	0.780	0.442	0.340	0.168	0.390	0.250	0.082
W-sr	0.825	0.735	0.499	0.275	0.184	0.067	0.372	0.254	0.096

Table 3: Empirical frequencies of rejection based on 2500 replications, $T = 400$, $B = 199$

In the hypothetical model H_{0B}, the innovations are independent and standard normally distributed; under alternative H_{1B}, innovations are independent with $\varepsilon_t \sim t(7)$.

The first part of Table 4 shows again that all tests maintain their theoretical level. For the alternative H_{0B}, the tests based on squared residuals are preferable, all with similar power. In the other group of tests, the most powerful tests are again the ones based on the Epps-Pulley and Watson statistic.

In the last trial we used once more a GARCH(1,1) model, but with parameter vector $(\beta_0, \beta_1, \gamma_1) = (0.1, 0.15, 0.8)$, which is similar to the fitted values in the real data example in Section 5 (see Table 8). In particular, $\beta_1 + \gamma_1$ is quite close to 1. Again, in the hypothetical and alternative model H_{0C} and H_{1C}, innovations are independent

	$H_{0,B}$		$H_{1,B}$		$H_{0,C}$		$H_{1,C}$	
	0.1	0.05	0.1	0.05	0.1	0.05	0.1	0.05
EP	0.095	0.046	0.338	0.207	0.112	0.053	0.515	0.326
CVM	0.092	0.048	0.218	0.137	0.102	0.052	0.256	0.130
HKT-CM	0.090	0.045	0.246	0.167	0.104	0.051	0.278	0.174
HKT-NCM	0.089	0.043	0.229	0.153	0.103	0.054	0.245	0.150
KS	0.096	0.044	0.249	0.156	0.106	0.053	0.296	0.174
HKT-KS	0.098	0.045	0.241	0.158	0.106	0.060	0.256	0.147
AD	0.091	0.049	0.253	0.155	0.101	0.057	0.335	0.176
W	0.092	0.045	0.436	0.303	0.104	0.060	0.658	0.532
LT-sr	0.089	0.050	0.567	0.432	0.100	0.053	0.840	0.768
CVM-sr	0.096	0.048	0.566	0.443	0.100	0.050	0.808	0.736
HKT-CM-sr	0.112	0.060	0.606	0.478	0.108	0.055	0.862	0.806
HKT-NCM-sr	0.116	0.068	0.585	0.460	0.104	0.058	0.830	0.765
KS-sr	0.100	0.045	0.532	0.404	0.105	0.054	0.769	0.686
HKT-KS-sr	0.100	0.047	0.548	0.419	0.103	0.058	0.862	0.777
AD-sr	0.097	0.050	0.544	0.407	0.101	0.053	0.804	0.729
W-sr	0.102	0.048	0.436	0.326	0.103	0.053	0.634	0.516

Table 4: Empirical frequencies of rejection based on 2500 replications, $T = 400$, $B = 199$

with $\varepsilon_t \sim N(0,1)$ and $\varepsilon_t \sim t(7)$, respectively.

For the given choice of parameter values, it is possible that $\hat{\beta}_1 + \hat{\gamma}_1 > 1$. If this was the case in the first estimation step, we rescaled $\hat{\beta}_1$ and $\hat{\gamma}_1$ such that $\hat{\beta}_1 + \hat{\gamma}_1 = 0.99$ in order to enable bootstrap simulation. However, this only happened in 5 out of 2500 cases for the normal and in about 1% of cases for the t distribution. The second part of Table 4 shows that the empirical level is slightly above the theoretical level and that powers against the t_7 distribution are generally higher than in the previous simulation. However, the ranking of the different tests remains more or less unchanged.

3.2. Testing for symmetry

To test the symmetry null hypothesis that ε_t and $-\varepsilon_t$ have the same distribution, we apply a similar bootstrap scheme as in Subsection 3.1 with the following modifications: in the first step, parameter estimation is always done by (Gaussian) QML method, in steps (ii) and (v) $\Phi_{T,w}$ is replaced by $\Sigma_{T,w}$, while step (iii) is modified as follows:
(iii') Define the bootstrap residuals

$$\hat{\varepsilon}_t^* = v_t \hat{\varepsilon}_t,$$

where $v_t,\ t = m+1, ..., T$, are i.i.d. observations with $\mathrm{P}(v_t = 1) = \mathrm{P}(v_t = -1) = 1/2$, and $v_{m+1}, \ldots, v_T$ are also independent of $\hat{\varepsilon}_{m+1}, \ldots, \hat{\varepsilon}_T$.
Similar bootstrap schemes were employed by Henze et al. (2003) for random vectors with an unspecified center and by Neumeyer et al. (2005) and Neumeyer and Dette (2007) in testing for symmetry in the context of linear and nonparametric regression.

To assess the actual level and the power of the test for symmetry based on Σ_T, a simulation study was performed for sample size $T = 400$, $B = 199$ bootstrap replications and the following symmetric and skewed distributions:

Symmetric: Normal, Laplace, t_5, t_7.

Skewed: Skew Normal, Skew Laplace, Skew t_5, Skew t_7.

In each case, the density of the skewed distribution, indexed by a scalar $\gamma \in (0, \infty)$, is generated from the density f of the pertaining unimodal and symmetric distribution by

$$f_\gamma(t) = \frac{2}{\gamma + 1/\gamma} \left\{ f(t/\gamma)\, \mathbf{1}_{[0,\infty)}(t) + f(\gamma t)\, \mathbf{1}_{(-\infty,0)}(t) \right\}.$$

Here, f_γ is symmetric for $\gamma = 1$, and skewed to the right (left) for $\gamma > 1$ ($\gamma < 1$), respectively. Furthermore, we get $f_\gamma(t) = f_{1/\gamma}(-t)$; therefore, it is sufficient to restrict ourself to values $\gamma > 1$. This method has been proposed by Fernández and Steel (1998). The skewed distributions mentioned above are available in the R package fGarch (Wuertz and Chalabi (2009)).

As data generating processes, we used the same ARCH(2) and GARCH(1,1) process

	$N(0,1)$	$Lap(1/\sqrt{2})$	t_5	t_7
ARCH(2)	0.050	0.051	0.043	0.048
GARCH(1,1)	0.056	0.054	0.045	0.054

Table 5: Empirical frequencies of rejection for the symmetry test based on 2500 replications; $T = 400$, $B = 199$, $\alpha = 0.05$

		Skew Normal			Skew Laplace		
	γ :	1.3	1.9	5.0	1.1	1.2	1.3
	Skewness:	0.39	0.76	0.97	0.40	0.74	1.01
ARCH(2)		0.42	0.96	1.00	0.25	0.72	0.94
GARCH(1,1)		0.42	0.96	1.00	0.24	0.73	0.95
		Skew t_5			Skew t_7		
	γ :	1.1	1.2	1.3	1.15	1.3	1.7
	Skewness:	0.43	0.78	1.08	0.41	0.74	1.05
ARCH(2)		0.16	0.42	0.72	0.23	0.65	0.94
GARCH(1,1)		0.16	0.44	0.72	0.23	0.63	0.92

Table 6: Empirical frequencies of rejection for the symmetry test based on 2500 replications; $T = 400$, $B = 199$, $\alpha = 0.05$

as in Subsection 3.1. For each underlying distribution and each process, the procedure described above was replicated 2 500 times. Results for $\alpha = 0.05$ are given in Table 5 (symmetric distributions) and Table 6 (skewed distributions). Approximate values of the usual moment based skewness are given in the third and 8th row of Table 6. The values of γ were chosen in such a way that the skewness has comparable values across the different distributions.

The results reveal that the theoretical level is maintained well for all symmetric distributions. Concerning the skewed alternatives, power increases with skewness. Moreover, we point out that power varies between different families of innovation dis-

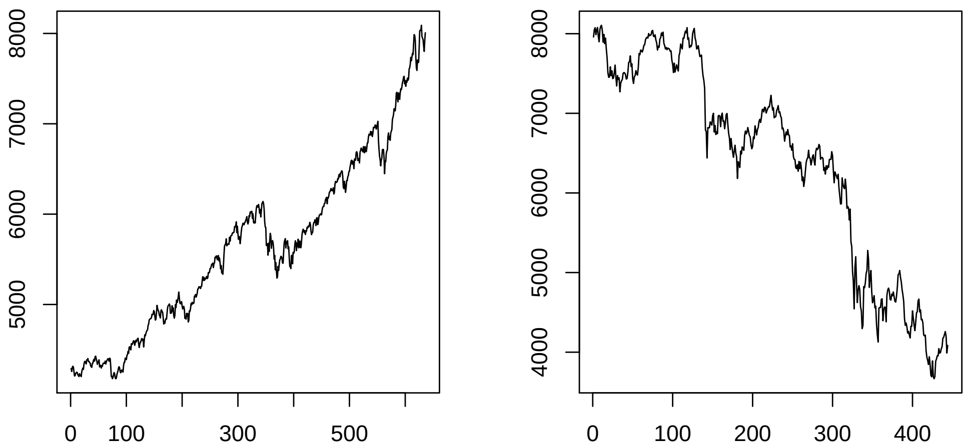

Figure 1: Time series of daily closing prices of the DAX index for period from 1 January 2005 to 30 June 2007 (left) and 1 July 2007 to 31 March 2009 (right)

tributions with common skewness, while, at the same time, it remains invariant to the data generating process under the same error distribution.

4. Application to DAX log returns

As an example, we apply the goodness–of–fit tests to different time series of log returns of the German stock market index DAX, consisting of the 30 major German companies trading on the Frankfurt Stock Exchange.

We have been interested in differences in the stochastic behaviour of the log returns between a period of economic growth and, consequently, a strong increase in the stock index, and a period where the index sharply declines. Since the beginning of the subprime mortgage crisis is commonly dated to July 2007, we selected the daily closing prices of the DAX index for the period from 1 January 2005 to 30 June 2007 as a first time series with 636 observations (further denoted by DAX05) and the period from 1 July 2007 until the end of the first quarter 2009 as a second time series (444 observations, denoted by DAX07). The two series, which are available online under `http://www.markt-daten.de/daten/daten.htm`, are shown in Figure 1. The visual appearance of the log returns of the two time series is quite different, as Figure 2 demonstrates.

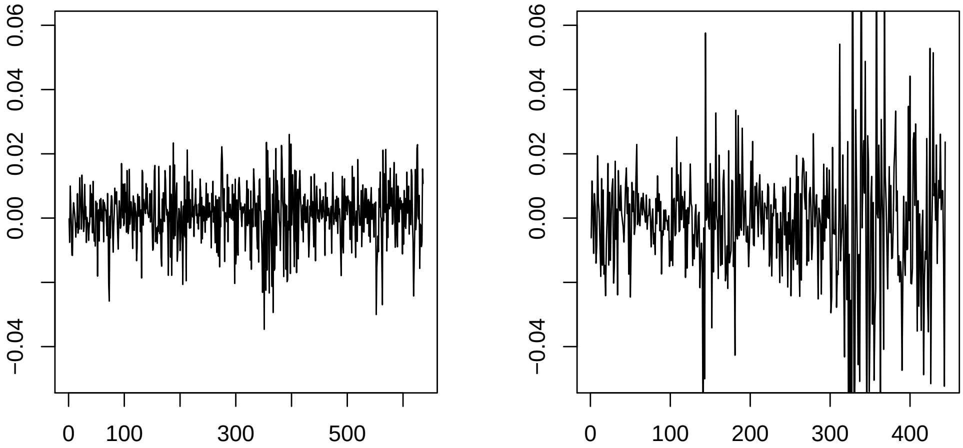

Figure 2: Time series of log returns of DAX05 (left) and DAX07 (right) time–series

	Sample size	Median	Mean	SD	Skewness	Kurtosis
DAX91	1786	0.091	0.068	1.051	-0.550	5.91
DAX05	636	0.146	0.098	0.885	-0.397	1.00
DAX07	444	-0.054	-0.152	2.077	0.412	5.64

Table 7: Statistical description of log returns of the three time–series

For comparison, we also considered DAX closing prices from 1991 to 1998. During this period, the DAX index increased from 1629 to 5474 points. This long time series is available as data set EuStockMarkets in the statistics software R (R (2008)) which we also used for all computations, together with the packages fGarch (Wuertz and Chalabi (2009)), TSA (Chan, 2008) and VGAM (Yee (2009)).

Contrary to the help page, EuStockMarkets includes holidays: about 4% of the DAX time series are successive equal prices. After removing these values, the remaining series, denoted by DAX91, has 1786 observations. Table 7 presents a statistical description of the three daily return series. The entries in the last column are the excess kurtosis; a value of zero corresponds to normality.

	$\hat{\mu}$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\gamma}_1$	$\hat{\beta}_1 + \hat{\gamma}_1$
DAX91	6.78e-04	5.72e-06	7.55e-02	8.74e-01	0.949
DAX05	1.18e-03	6.25e-06	9.47e-02	8.24e-01	0.919
DAX07	-9.89e-05	5.05e-06	1.43e-01	8.59e-01	1.001

Table 8: Coefficients of a GARCH(1,1) model fitted to the three time–series

As expected, mean and median of the first two time series are positive, whereas mean and median of the last series are negative. The empirical distributions of log returns of the DAX and other indices are often found to be too leptokurtic (compared to the normal) and skewed to the left (for example, Laplante et al. (2008)). Indeed, the kurtosis is positive for all time series, but with a noticeable difference between the second and the other two data sets. Furthermore, the third series is skewed to the right.

Since the means of the return series deviate form zero, we fitted the following GARCH(1,1) models to the three return series:

$$y_t = \mu + c_t \varepsilon_t, \quad c_t^2 = \beta_0 + \beta_1 (y_{t-1} - \mu)^2 + \gamma_1 c_{t-1}^2.$$

As it has been observed in the past, a GARCH(1,1) model seems to be a suitable model for DAX log returns. Sapusek (2004), for example, compared several models for long-term DAX series and identified a GARCH(1,1) model for the variance as the most adequate; see also Angelidisa et al. (2004) or Laplante et al. (2008). The autocorrelation functions of the squared residuals for the DAX05 and DAX07 data displayed in Figure 3 point in this direction: whereas the squared log returns are highly correlated, this is no longer the case for the squared residuals.

The coefficients of the GARCH(1,1) model fitted to each of the three time series are comparable, with a higher value of $\hat{\beta}_1$ for DAX07 compared to the other two series (cf. Table 8).

The entries in the last column of Table 8 show the values of $\hat{\beta}_1 + \hat{\gamma}_1$. The condition $\beta_1 + \gamma_1 < 1$ ensures that the process is covariance stationary. The parameter estimates

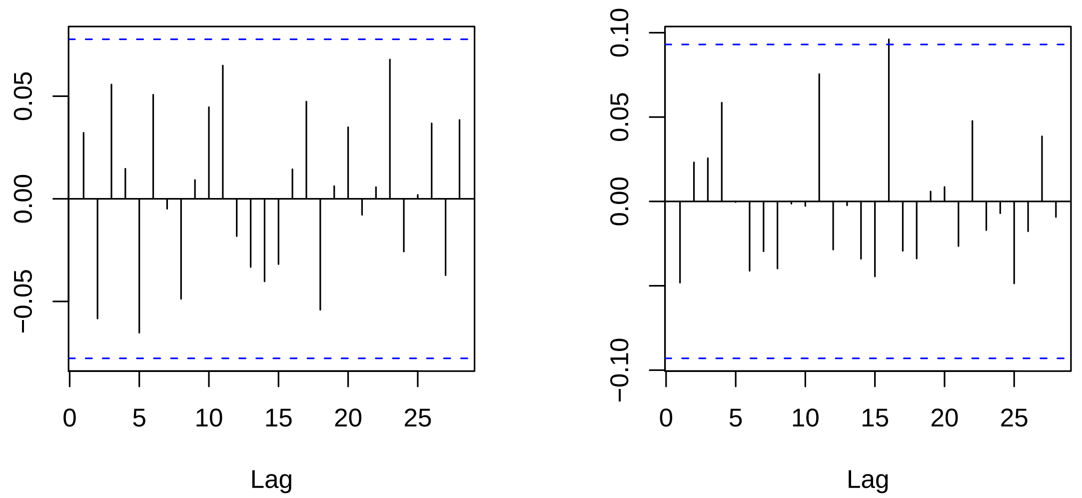

Figure 3: Autocorrelation function of squared residuals of a GARCH(1,1)-model fitted to the DAX05 (left) and DAX07 (right) time–series

of β_1 and γ_1 for the DAX07 series violate this condition. We have tried several return series with slightly differing periods; the sum $\hat{\beta}_1+\hat{\gamma}_1$ was sometimes greater, sometimes less than 1, but in all cases very near to this bound.

We tested the hypothesis of normality of the innovations using all tests of Section 5 using $B = 1999$ bootstrap samples. For the DAX07 series, we rescaled $\hat{\beta}_1$ and $\hat{\gamma}_1$ in such a way that $\hat{\beta}_1 + \hat{\gamma}_1 = 0.99$, as described at the end of Subsection 3.1; in practice, one would probably desist from using a model violating the stationarity condition. For the three time series, all p–values are very close to zero. The innovations of the processes definitely do not follow a normal distribution.

Moreover, we considered the Laplace distribution (with scale parameter $1/\sqrt{2}$) and t–distribution with $k = 5, 6, 7$ degrees of freedom (each scaled to unit variance) as hypothetical distributions. The latter was proposed by Bollerslev (1987) as conditional error distribution. Here, we made use of the EDF tests of Cramér–von Mises, Kolmogorov–Smirnov, Anderson–Darling and Watson based on the non-squared residuals. Additionally, we applied the Epps-Pulley test (again based on the non-squared residuals) using the appropriate hypothetical characteristic functions $\varphi_L(u) = (1+u^2/2)^{-1}$ for the

	$N(0,1)$	$Lap(1/\sqrt{2})$	t_5	t_6	t_7
EP	0.00	0.00	0.01	0.04	0.09
CM	0.00	0.00	0.18	0.38	0.49
KS	0.00	0.00	0.28	0.54	0.6
AD	0.00	0.00	0.12	0.34	0.48
WA	0.00	0.00	0.19	0.48	0.61
EP	0.00	0.00	0.03	0.04	0.04
CM	0.00	0.05	0.02	0.01	0.00
KS	0.00	0.02	0.02	0.04	0.04
AD	0.00	0.01	0.00	0.00	0.00
WA	0.00	0.02	0.04	0.02	0.01
EP	0.00	0.00	0.09	0.11	0.11
CM	0.00	0.00	0.03	0.03	0.02
KS	0.00	0.00	0.12	0.13	0.07
AD	0.00	0.00	0.02	0.02	0.01
WA	0.00	0.01	0.23	0.18	0.16

Table 9: p–values of tests for different hypothetical error distributions; upper part: DAX91, middle: DAX05, lower part: DAX07

Laplace distribution and $\varphi_k(u) = \varphi_{t_k}(u\sqrt{(k-2)/k})$ for the rescaled t–distribution with k degrees of freedom. The corresponding characteristic function of the t_k-distribution is given by (see Hurst (1995))

$$\varphi_{t_k}(u) = \frac{K_{k/2}(\sqrt{k}|u|)(\sqrt{k}|u|)^{k/2}}{\Gamma(k/2)2^{k/2-1}},$$

where $K_\lambda(\cdot)$ is the modified Bessel function of the third kind with index λ and $\Gamma(\cdot)$ is the Gamma function.

The results for the time three series are given in Table 9. Again, the entries are based on $B = 1999$ bootstrap samples, and estimation was done by the method of maximum likelihood. In all cases, the sum $\hat{\beta}_1 + \hat{\gamma}_1$ was less than 1.

For the Laplace distribution, the p–values vary between 0 and 0.05, but are close

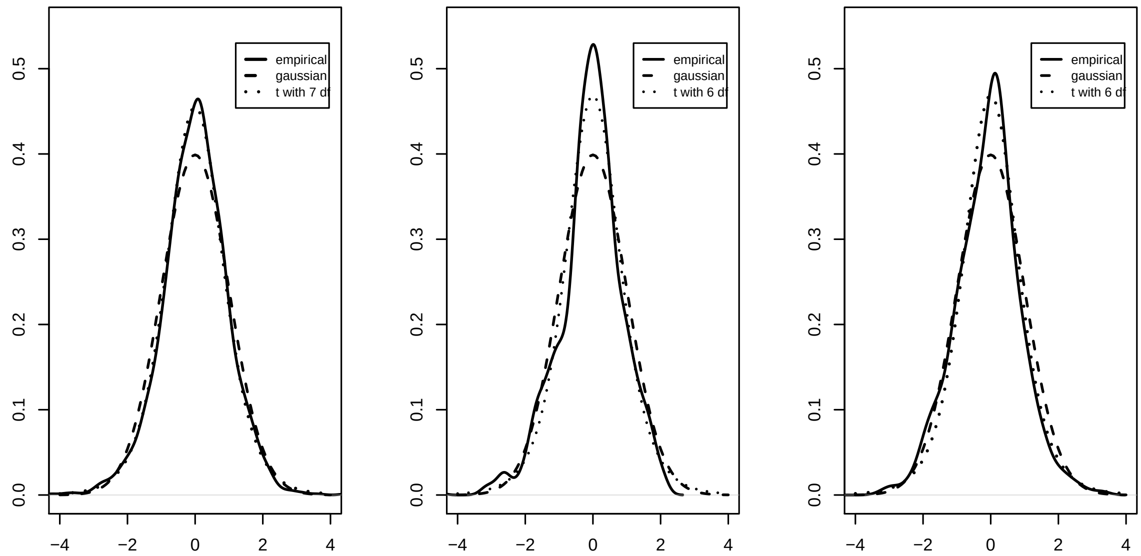

Figure 4: Empirical and theoretical density functions of residuals of a GARCH(1,1)-model fitted to the DAX91 (left), DAX05 (middle) and DAX07 (right) time–series

to zero in most cases. Concerning the t–distribution, the results depend on the data set. Looking at the DAX91 data, the different t–distributions are not rejected on the 5% level by the EDF tests; the test based on the ECF does not reject the t–distribution with 7 degrees of freedom. For the DAX05 series, the p–values are between 0 and 0.05. The tests of Anderson–Darling and Cramér–von Mises yield rather low p–values for the DAX07 series, contrary to the remaining three tests.

A visual inspection of the distribution of the residuals corroborates the results of the formal goodness of fit tests. Figure 4 shows plots of the empirical densities together with the theoretical densities of the normal and the t distribution with 7 (DAX91) and 6 (DAX05, DAX07) degrees of freedom.

The empirical density of the DAX91 series is fairly symmetric; the fit of the t_7 distribution is excellent. Considering the DAX07 series, the t_6 density fits quite well, whereas the normal density is not sufficiently peaked. The symmetry test yields p-values of 0.37 and 0.18, respectively. The distribution of the DAX05 residuals can not be modeled satisfactorily by the distributions at hand. It looks somewhat asymmetric; however, this is not confirmed by the symmetry test resulting in a p-value of 0.17.

To sum up the results of this section, there exist differences between the time series: the parameters of the GARCH process are similar, but only for the DAX07 series the sum $\hat{\beta}_1 + \hat{\gamma}_1$ is very close to 1, regardless of the hypothetical distribution. Moreover, there are pronounced differences between the shapes of the error distributions of the three data sets.

5. Conclusion

GARCH models with Gaussian innovations may be viewed as a benchmark in the attempt to adequately describe some so–called stylized facts that have been observed across a wide range of financial data sets, such as volatility clustering and heavy tails. However, incorrect specification of the innovation distribution may lead to sizeable loss of efficiency of the corresponding estimators. Also, from the point of view of a financial analyst, this misspecification could imply invalid risk determination, inaccurately priced options and wrong assessment of Value–at–Risk (VaR).

In this paper we examine the behaviour of tests of fit for the hypothesis of normality of innovations in GARCH models. The procedures are natural extensions of well-known tests for normality, which include classical goodness–of–fit tests based on the empirical distribution function as well as more recent tests utilizing empirical transforms. One of these transforms is also used for the construction of a specification test for symmetry of the innovation distribution. All methods incorporate an intermediate estimation step, and use either the resulting residuals or the corresponding squared residuals. Despite the fact that there is no clear preference of one version over the other (squared or non–squared residuals), it appears that in the case of normality the Watson test and the test of Epps–Pulley are most powerful. The symmetry test also behaves well.

Although the simulation results presented are restricted to normality tests, our applications on real data suggest that the proposed methods can be applied equally well to some popular alternatives to the Gaussian GARCH. Such GARCH models could include not only the location–scale models considered herein, but also Student's t–innovations with unknown degrees of freedom (possibly non–integer), or the GARCH–

stable model of Mittnik et al. (1998) in which innovations follow a symmetric stable distribution with unknown characteristic exponent. There is an added complication of course in the latter cases of distributions, since, on top of the estimation of the GARCH parameters, there is an extra shape parameter which also needs to be estimated from the data prior to testing goodness–of–fit for the innovation distribution.

Acknowledgement. Part of this research was carried out while the last author was visiting the University of Karlsruhe (now Karlsruhe Institute of Technology (KIT)). SGM wishes to sincerely thank the Institute of Stochastics for the hospitality and financial support.

References

Angelidisa, T., Benosa, A., Degiannakis, S., 2004. The use of GARCH models in VaR estimation. Statistical Methodology 1 (1-2), 105–128.

Bellini, F., Bottolo, L., 2009. Misspecification and domain issues in fitting GARCH(1,1) models: A Monte Carlo investigation. Commun. Statist. Simul. Comput. 38 (1), 31–45.

Berkes, I., Horváth, L., 2004. The efficiency of estimators of the parameters in GARCH processes. Ann. Statist. 32 (2), 633–655.

Berkes, I., Horváth, L., Kokoszka, P., 2003. GARCH processes: Structure and estimation. Bernoulli 9 (2), 201–228.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroscedasticity. Journal of Econometrics 31 (3), 307–327.

Bollerslev, T., 1987. A conditionally heteroscedastic time series model for speculative prices and rates of return. The Review of Economics and Statistics 69 (3), 542–547.

Bougerol, P., Picard, N., 1992a. Stationarity of GARCH processes and some nonnegative time series. J. Econometr. 52 (1-2), 115–127.

Bougerol, P., Picard, N., 1992b. Strict stationarity of generalized autoregressive processes. The Annals of Probability 20 (4), 1714–1730.

Chan, K., 2008. TSA: Time series analysis. R package version 0.97. URL=http://www.stat.uiowa.edu/ kchan/TSA.htm.

Csörgő, S., 1981. Limit behaviour of the empirical characteristic function. Ann. Probab. 9 (1), 130–144.

Csörgő, S., Totik, V., 1983. On how long an interval is the empirical characteristic function uniformly consistent. Acta Sci. Math. (Szeged) 45, 141–149.

Curto, J., Pinto, J., Tavares, G., 2009. Modeling stock markets' volatility using GARCH models with normal, Student's t and stable Paretian distributions. Statist. Papers 50 (2), 311–321.

D'Agostino, R., Stephens, M., 1986. Goodness–of–fit techniques. Marcel Dekker, Inc.

Engle, R., González-Rivera, G., 1991. Semi–parametric ARCH models. J. Busin. Econom. Statist. 9 (4), 345–359.

Epps, T., 2005. Tests for location–scale families based on the empirical characteristic function. Metrika 62 (1), 99–114.

Epps, T., Pulley, L., 1983. A test for normality based on the empirical characteristic function. Biometrika 70 (3), 723–726.

Fernández, C., Steel, M., 1998. On Bayesian modeling of fat tails and skewness. Journal of the American Statistical Association 93, 359–371.

Feuerverger, A., Mureika, R., 1977. The empirical characteristic function and its applications. Annals of Statistics 5 (1), 88–97.

Forsberg, L., 2002. On the normal inverse Gaussian distribution in modeling volatility in the financial markets. PhD thesis, Uppsala University.

Francq, C., Zakoïan, J., 2004. Maximum likelihood estimation of pure GARCH and ARMA–GARCH processes. Bernoulli 10 (4), 605–637.

Francq, C., Zakoïan, J., 2006. On efficient inference in GARCH processes. In: Bertail, P., Doukhan, P., Soulier, P. (Eds.), Lecture Notes in Statistics No. 187: Dependence in probability and statistics. Spinger, New York, pp. 305–327.

Giraitis, L., Robinson, P., Surgailis, D., 2000. A model for long memory conditional heteroscedasticity. The Annals of Applied Probability 10 (3), 1002–1024.

González-Rivera, G., Drost, F., 1999. Efficiency comparisons of maximum–likelihood–based estimators in GARCH models. J. Econometr. 93 (1), 93–111.

Granger, C., Ding, Z., 1995. Some properties of absolute return, an alternative measure of risk. Ann. Econom. Statist. 40, 67–91.

Hansen, P., Lunde, A., 2005. A forecast comparison of volatility models: Does anything beat the GARCH(1,1)? J. Appl. Econometr. 20 (7), 8783–889.

Henze, N., Klar, B., Meintanis, S., 2003. Invariant tests for symmetry about an unspecified point based on the empirical characteristic function. J. Multivar. Anal. 87 (2), 275–297.

Henze, N., Wagner, T., 1997. A new approach to the BHEP tests for multivariate normality. J. Multivar. Anal. 62 (1), 1–23.

Horváth, L., Kokoszka, P., Teyssiére, G., 2004. Bootstrap misspecification tests for ARCH based on the empirical process of squared residuals. J. Statist. Comput. Simul. 74 (7), 469–485.

Horváth, L., Zitikis, R., 2006. Testing goodness of fit based on densities of GARCH innovations. Econometr. Theor. 22 (3), 457–482.

Huang, D., Wang, H., Yao, Q., 2008. Estimating GARCH models: When to use what? Econometrics Journal 11 (1), 27–38.

Hurst, S., 1995. The characteristic function of the student t distribution. Statistics Research Report SRR044–95, Australian National University.

Jarque, C., Bera, A., 1987. A test for normality of observations and regression residuals. Inter. Statist. Rev. 55 (2), 163–172.

Koul, H., Ling, S., 2006. Fitting an error distribution in some heteroscedastic time series models. Ann. Statist. 34 (2), 994–1012.

Kulperger, R., Yu, H., 2005. High moment partial sum processes of residuals in GARCH models and their applications. Ann. Statist. 33 (5), 2395–2422.

Laplante, J., Desrochers, J., Préfontaine, J., 2008. The GARCH(1,1) model as a risk predictor for international portfolios. International Business & Economics Research Journal 7 (11), 23–34.

Marcus, M., 1981. Weak convergence of the empirical characteristic function. Ann. Probab. 9 (2), 194–201.

Matsui, M., Takemura, A., 2007. Goodness–of–fit tests for symmetric stable distributions–empirical characteristic function approach. TEST 17 (3), 546–566.

Meintanis, S., 2004. A class of omnibus tests for the Laplace distribution based on the empirical characteristic function. Commun. Statist. Theor. Meth. 4 (33), 925–948.

Meintanis, S., Swanepoel, J., 2007. Bootstrap goodness–of–fit tests with estimated parameters based on empirical transforms. Statist. Probab. Lett. 77 (10), 1004–1013.

Mittnik, S., Paolella, M., 2003. Prediction of financial downside–risk with heavy tailed conditional distributions. In: Rachev, S. (Ed.), Handbook of Heavy–Tailed Distributions in Finance. Elsevier.

Mittnik, S., Rachev, S., Paolella, M., 1998. Stable Paretian modeling in finance. Birkhauser Boston Inc., pp. 97–110.

Nelson, D., 1991. Conditional heteroscedasticity in asset return: a new approach. Econometrica 59 (2), 347–370.

Neumeyer, N., Dette, H., 2007. Testing for symmetric error distribution in nonparametric regression models. Statist. Sinica 17, 775–795.

Neumeyer, N., Dette, H., Nagel, E., 2005. A note on testing symmetry of the error distribution in linear regression models. J. Nonparametr. Statist. 17 (6), 697–715.

Newey, W., Steigerwald, D., 1997. Asymptotic bias for quasi–maximum likelihood in conditional heteroscedastic models. Econometrica 65 (3), 587–6000.

Ngatchou-Wandji, J., 2009. Testing symmetry of the error distribution in nonlinear heteroskedastic models. Commun. Statist. Theor. Meth. 38 (9), 1465–1485.

R, 2008. R: A language and environment for statistical computing. R foundation for statistical computing. URL=http://www.R-project.org.

Sapusek, A., 2004. Volatility clustering in German stock returns. Schmidt, R. and Gramlich, D., pp. 81–92.

Wuertz, D., Chalabi, Y., 2009. fGarch: Rmetrics - autoregressive conditional heteroskedastic modelling. R package version 2100.79. URL=http://CRAN.R-project.org/package=fGarch.

Yee, T., 2009. VGAM: Vector generalized linear and additive models. R package version 0.7–8. URL=http://www.stat.auckland.ac.nz/ yee/VGAM.