Testing for GARCH effects with quasi-likelihood ratios

Richard Luger[1]

Department of Risk Management and Insurance, Georgia State University, Atlanta, GA 30303, USA; e-mail: rluger@gsu.edu

A procedure is developed to test whether conditional variances are constant over time in the context of GARCH models with possible GARCH-in-mean effects. The approach is based on the quasi-likelihood function, leaving the true distribution of model disturbances parametrically unspecified. The presence of possible nuisance parameters in the conditional mean is dealt with by using a pivotal bound and Monte Carlo resampling techniques to obtain a level-exact test procedure. Simulation experiments reveal that the permutation-based, quasi-likelihood ratio test has very attractive power properties in comparison to omnibus Lagrange multiplier tests. An empirical application of the new procedure finds overwhelming evidence of GARCH effects in Fama-French portfolio returns, even when conditioning on the market risk factor.

KEYWORDS: Conditional heteroskedasticity; GARCH-M; Quasi-maximum likelihood estimation; Distribution-free Monte Carlo test; Exact inference

1 Introduction

The autoregressive conditional heteroskedasticity (ARCH) model of Engle (1982) and the generalized ARCH (GARCH) model of Bollerslev (1986) are very popular specifications for modeling the time variation in the conditional variance of financial series. For instance, Lee and Hansen (1994) refer to the GARCH(1,1) model as the "workhorse of the industry," which speaks to the popularity of this specification. GARCH models are extremely useful for financial risk modeling because they are relatively easy to estimate compared to, say, stochastic volatility models and they can account for many well-known features such as the volatility clustering of financial returns.

[1]I would like to thank Lynda Khalaf, Don Andrews, Russell Davidson, and Eric Renault for useful remarks made on earlier versions of this paper. The suggestions of an anonymous referee are also gratefully acknowledged.

The estimation of GARCH models is often performed by the method of quasi- (or pseudo-) maximum likelihood, which proceeds by maximizing the Gaussian likelihood function even though the true innovation distribution may not be normal. Gouriéroux, Monfort, and Trognon (1984) and Bollerslev and Wooldridge (1992) show that such estimators are nevertheless consistent under certain regularity conditions. The consistency and asymptotic distribution of the quasi-maximum likelihood estimator of GARCH model parameters is shown in Weiss (1986), Lee and Hansen (1994), Lumsdaine (1996), Berkes, Horváth, and Kokoszka (2003), Jensen and Rahbek (2004a,b), and Francq and Zakoïan (2004, 2007).

The key regularity conditions for the asymptotic distribution of the quasi-maximum likelihood estimator restrict the innovation distribution to have certain finite moments and they also restrict the true parameter values to lie in the interior of the parameter space. The latter condition is violated under the null hypothesis of conditional homoskedasticity. Indeed, one of the boundaries of the parameter space for GARCH models restricts the parameters to be non-negative in order to ensure that the conditional variance process itself is non-negative. (Nelson and Cao (1992) give weaker conditions, but these are generally not explicit.) In the absence of a time-varying conditional variance, the ARCH parameters are zero and this in turn implies that the GARCH parameters are not identified. When this happens the quasi-likelihood ratio (QLR) statistic does not have the usual asymptotic χ^2 distribution, but rather follows a non-standard and quite complicated distribution involving nuisance parameters (Andrews, 2001; Francq and Zakoïan, 2010, Ch. 8).

The QLR approach is not the only way to test for GARCH effects. In fact, the very popular Engle (1982) test is based on the Lagrange multiplier (LM) approach and retains the usual asymptotic χ^2 distribution under the null hypothesis of conditional homoskedasticity. When the alternative hypothesis of a GARCH process holds true, some of the ARCH parameters are strictly positive. By taking into account this one-sided nature of the GARCH alternative, Lee and King (1993) and Demos and Sentana (1998) propose different LM tests which yield more power than Engle's two-sided LM test. All these LM tests are omnibus in the sense that they detect departures from conditional homoskedasticity in general GARCH directions. However, they are not formulated to detect any specific GARCH alternative.

Given the widespread use of GARCH specifications in financial risk modeling, it is clearly of interest to have a valid and reliable QLR test of conditional homoskedasticity that can easily be applied in the context of the specific GARCH model under consideration. Note that, by its very definition, the QLR statistic yields a directional (i.e. one-sided) test. Andrews (2001) derives an asymptotic QLR test of conditional homoskedasticity by using cones to approximate the parameter space, but only against the GARCH(1,1) model of Bollerslev (1986). Andrews' test proceeds by finding the supremum of a QLR statistic over the possible values for the GARCH parameter, which is unidentified under the null hypothesis. In general, the asymptotic null distribution of that statistic is nuisance-parameter dependent. Beg, Silvapulle, and Silvapulle (2001) also propose one-sided tests of conditional homoskedasticity that deal with the fact that some parameters are present only under the alternative hypothesis, but their approach is developed only for ARCH models with normally distributed innovations.

In this paper, a simulation-based procedure is developed to test the null hypothesis of conditional homoskedasticity in the context of specified GARCH models. The conditional variance specification may be any GARCH model of the user's choice, and the conditional mean specification allows for the GARCH-in-mean effect of Engle, Lilien, and Robins (1987) and possible exogenous explanatory variables (with parameters $\boldsymbol{\beta}$). A very popular GARCH model in applied work is the symmetric GARCH specification of Bollerslev (1986), where positive and negative past returns have the same effect on the conditional variance. Yet the empirical literature has documented cases where negative returns tend to be followed by larger increases in volatility than do equally large positive ones. This stylized fact is often called the leverage effect because a negative return implies a drop in equity value, which in turn implies that the firm becomes more highly leveraged and hence more risky as long as its debt level stays constant. The asymmetric GARCH model of Glosten, Jagannathan, and Runkle (1993) is another popular specification that can capture this leverage effect. Both these specifications are used here for illustration purposes in the simulation experiments and the empirical application.

The proposed approach is based on the quasi-likelihood function, leaving the true distribution of model innovations parametrically unspecified. Indeed, the only assumption made is the usual one that the innovations are independently and identically distributed (i.i.d.) according to a mean zero and unit variance distribution. Among several other regularity conditions, the asymptotic theory for the QLR presented in Andrews (2001) and Francq and Zakoïan (2010, Ch. 8) assumes that the innovations have a finite moment of order four. This assumption may be questionable when dealing with financial returns which are notorious for their heavy tails (cf. Politis, 2004). In sharp contrast, the innovation distribution here is not restricted to have finite moments beyond order two.

A permutation principle is first established to test the null hypothesis of constant conditional variances when the values of $\boldsymbol{\beta}$ are specified. This permutation principle yields an equally likely property for the QLR statistic and this forms the basis for an exact procedure using the technique of Monte Carlo (MC) tests (Barnard, 1963; Birnbaum, 1974; Dwass, 1957). Unlike the MC tests for ARCH-type heteroskedasticity of Dufour et al. (2004) whose exactness rests on a maintained distributional assumption (e.g. normal innovations), the procedure proposed here is distribution-free. A similar approach is developed in Luger (2006) to test the specification of regression models against non-nested alternatives.

The permutation principle also paves the way for a three-step procedure to deal with situations where the values of $\boldsymbol{\beta}$ are left unspecified, thereby becoming nuisance parameters in the present context. The first step simply consists of the MC test performed at the quasi-maximum likelihood estimate of $\boldsymbol{\beta}$ obtained under conditional homoskedasticity. If this step is not significant, then conditional homoskedasticity is immediately accepted. Otherwise the procedure moves on to the second step, which consists of an MC test based on a pivotal bound to the finite-sample null distribution of the QLR statistic. Following the approach in Dufour and Khalaf (2002, 2003), it is shown that a rejection by the bounds test implies the rejection of conditional homoskedasticity in favor of the maintained GARCH alternative. Finally a non-rejection by the bounds MC test leads to step three which consists of maximizing the MC p-value, following a *minimax* argument to deal with the presence of nuisance parameters. This test procedure is guaranteed to control the overall probability of a Type I error in finite samples, even though the GARCH-in-mean parameter (in the conditional mean) and the GARCH parameters themselves are not identified under the null hypothesis. See also Khalaf, Saphores, and Bilodeau (2003) for a similar three-pronged testing strategy in a different context.

The paper proceeds as follows. Section 2 begins by presenting the statistical framework and then develops each step of the MC test procedure, summarized at the end of the section. Section 3 presents the results of simulation experiments comparing the size and power of the proposed procedure with the three popular LM tests mentioned above. Owing to its conservative nature in the presence of nuisance parameters, the MC test has relatively lower power than the LM tests for alternatives close to the null. But the simulation evidence reveals that the power of the MC test increases at a faster rate and surpasses the LM tests as the GARCH alternative moves away from the null. The power of the MC test also increases dramatically with the GARCH-in-mean parameter, whereas the LM tests barely respond. Section 4 presents an application of the GARCH tests to the well-known Fama and French (1993) portfolio returns data, which are routinely used in empirical asset pricing studies. The evidence of GARCH effects is much stronger on the basis on the new test procedure than on the basis of the LM tests. These results are interesting because the specification even controls for the expected returns of a benchmark market portfolio like in the traditional capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965). Section 5 concludes.

2 Test procedure

2.1 Framework and building blocks

Consider a financial asset with return (or excess return) y_t whose conditional mean specification allows for the GARCH-in-mean (GARCH-M) effect of Engle, Lilien, and Robins (1987). Specifically, the random variable y_t is assumed to be described as

$$y_t = \mu + \delta h(\sigma_t) + \boldsymbol{x}_t'\boldsymbol{\beta} + \varepsilon_t \tag{1}$$

$$\varepsilon_t = \sigma_t \eta_t \tag{2}$$

for $t = 1, ..., T$, where $\boldsymbol{x}_t$ is a vector of exogenous variables (i.e. that can be taken as fixed for statistical analysis); μ and δ are unknown parameters, and the vector $\boldsymbol{\beta}$ also contains unknown parameters. The potentially heteroskedastic error term ε_t in (2) comprises the conditional standard deviation (volatility) σ_t of y_t and an innovation term η_t, which is assumed to be i.i.d. according to a mean zero and unit variance distribution. Note that aside from this minimal assumption, the distribution of the innovations appearing in (2) is left completely unspecified. Indeed, the innovation distribution need not even have finite moments beyond order two. The presence of the function $h(\sigma_t)$ in the conditional mean of (1) can capture a potential trade-off between return volatility (i.e. risk) and expected return; commonly used forms include $h(\sigma_t) = \sigma_t^2$ and $h(\sigma_t) = \log(\sigma_t^2)$. Conditional mean equations of this form are widely used in empirical studies of time-varying risk premiums; well-known examples include Bollerslev (1987), French, Schwert, and Stambaugh (1987), Nelson (1991), Sentana (1995), and Hentschel (1995).

The conditional standard deviation σ_t in (2), if indeed time-varying, is assumed to evolve according to a GARCH process such that ε_t is covariance (or second-order) stationary. The most prominent example is the seminal GARCH(p, q) model of Bollerslev (1986) given by

$$\sigma_t^2 = \omega + \sum_{i=1}^{q} a_i \varepsilon_{t-i}^2 + \sum_{j=1}^{p} b_j \sigma_{t-j}^2 \tag{3}$$

where it is sufficient to impose $\omega > 0$, $a_i \geq 0$, $i = 1, ..., q$, and $b_j \geq 0$, $j = 1, ..., p$, in order to guarantee that $\sigma_t^2 > 0$ for all t. When $p = 0$, this model reduces to the ARCH(q) specification of Engle (1982). The parameters of (3) are further assumed to satisfy the constraint $\sum_{i=1}^{q} a_i + \sum_{j=1}^{p} b_j < 1$, which is sufficient to ensure that ε_t is covariance stationarity (Bollerslev, 1986; Bougerol and Picard, 1992; Nelson, 1990).

The inference problem considered here is to test the null hypothesis that $\sigma_t^2 = \sigma^2$, $t = 1, ..., T$, against the alternative that the error terms in (2) indeed follow a GARCH process. The value of σ^2 appearing in the null is the model-implied unconditional (stationary) variance. In the context of (3), this value is given by $\sigma^2 = \omega/(1 - \sum_{i=1}^{q} a_i - \sum_{j=1}^{p} b_j)$. The null hypothesis of conditional homoskedasticity can then be formulated in terms of the ARCH parameters $\boldsymbol{a} = (a_1, ..., a_q)'$ as $H_0 : \boldsymbol{a} = \mathbf{0}$. The GARCH alternative hypothesis is that at least one element of $\boldsymbol{a}$ is strictly greater than zero, and the elements of $\boldsymbol{b} = (b_1, ..., b_p)'$ are greater or equal to zero. Note that under H_0, the statistical identification of the GARCH parameters in $\boldsymbol{b}$ breaks down. Indeed, when $\boldsymbol{a} = \mathbf{0}$ the model in (3) can be written as

$$\sigma_t^2 = \frac{\omega}{1 - \sum_{j=1}^{p} b_j} + \sum_{j=1}^{p} b_j \left(\sigma_{t-j}^2 - \frac{\omega}{1 - \sum_{j=1}^{p} b_j} \right) = \vartheta + \sum_{j=1}^{p} b_j \left(\sigma_{t-j}^2 - \vartheta \right)$$

where $\vartheta = \omega/(1 - \sum_{j=1}^{p} b_j)$ corresponds to the unconditional variance under H_0. So starting with $\sigma_1^2 = ... = \sigma_p^2 = \vartheta$, all subsequent values are also given as $\sigma_t^2 = \vartheta$ and the parameters $b_1, ..., b_p$ disappear; see Andrews (2001) for further discussion of this identification problem. It is also easy to see that the risk premium parameter δ cannot be identified separately from μ in (1) under conditional homoskedasticity, since $h(\sigma)$ is a constant.

When the null hypothesis that $\sigma_t^2 = \sigma^2$, for $t = 1, ..., T$, is formulated as $\boldsymbol{a} = \mathbf{0}$, it assumes that $\sigma^2 = \vartheta$ which leaves the values of $b_1, ..., b_p$ free. If one does not wish to make this assumption, then the null of constant conditional variances could simply be restated as $\boldsymbol{a} = \mathbf{0}$ and $\boldsymbol{b} = \mathbf{0}$, so that the common variance becomes $\sigma^2 = \omega$. Of course the latter formulation assumes away the identification problem with $b_1, ..., b_p$ discussed above, but still leaves δ unidentified. A great advantage of the proposed test procedure is that it remains valid with both these formulations.

The model parameters to be estimated with the GARCH specification in (3) are collected in $\boldsymbol{\theta} = (\mu, \delta, \boldsymbol{\beta}', \omega, \boldsymbol{a}', \boldsymbol{b}')'$ and $\boldsymbol{\Theta}$ denotes the corresponding set of admissible values which ensures the positivity of σ_t^2 and the covariance stationarity of ε_t. Let $\boldsymbol{Y} = (y_1, ..., y_T)'$ represent the observed sample and denote by $m = \max(p, q)$ the number of observations lost for the initialization of the GARCH recursion. An immensely popular method of estimation for GARCH models is to proceed as if the innovation terms were normally distributed, even though they may not be in reality. This is the so-called method of Gaussian quasi- (or pseudo-) maximum likelihood; see Gouriéroux, Monfort, and Trognon (1984) and Bollerslev and Wooldridge (1992). The corresponding conditional (on the first m observations) quasi-log likelihood function is given by

$$QL(\boldsymbol{\theta}; \boldsymbol{Y}) = -\frac{1}{2} \sum_{t=m+1}^{T} \left(\log(2\pi) + \log\left(\sigma_t^2(\boldsymbol{\theta})\right) + \frac{\left(y_t - \mu_t(\boldsymbol{\theta})\right)^2}{\sigma_t^2(\boldsymbol{\theta})} \right) \tag{4}$$

where $\mu_t(\boldsymbol{\theta}) = \mu + \delta h(\sigma_t(\boldsymbol{\theta})) + \boldsymbol{x}_t'\boldsymbol{\beta}$ and the skedastic function $\sigma_t^2(\boldsymbol{\theta})$ is defined by (3). Two common ways of setting the initial values $\sigma_1^2, ..., \sigma_p^2$ in the GARCH recursion in (3) are as follows. The first is to use the model-implied unconditional variance, which of course depends on the parameter values. The second one proceeds somewhat less parametrically by using a first-pass estimate of the sample variance of ε_t obtained by the method of ordinary least squares (OLS). Either of these GARCH model initialization schemes may be used here. The quasi-maximum likelihood estimate (QMLE) obtained under the alternative hypothesis, $\hat{\boldsymbol{\theta}}_1$, satisfies

$$\hat{\boldsymbol{\theta}}_1(\boldsymbol{Y}) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmax}}\ QL(\boldsymbol{\theta}; \boldsymbol{Y})$$

where $QL(\boldsymbol{\theta}; \boldsymbol{Y})$ is the quasi-log likelihood function in (4).

It is well known that the conditional QMLEs of $c = \mu + \delta h(\sigma)$, $\boldsymbol{\beta}$, and σ^2 under conditional homoskedasticity can be obtained from a linear regression (cf. Hayashi, 2000, §1.5). Specifically, the H_0-restricted maximizers of (4) can be found by applying OLS to the model:

$$y_t = c + \boldsymbol{x}_t'\boldsymbol{\beta} + e_t \tag{5}$$

for $t = m+1, ..., T$. Let $\hat{c}_0$ and $\hat{\boldsymbol{\beta}}_0$ denote these estimates and define $\hat{\boldsymbol{\theta}}_0(\boldsymbol{Y}) = (\hat{c}_0, 0, \hat{\boldsymbol{\beta}}_0', \hat{\sigma}_0^2, \mathbf{0}', \mathbf{0}')'$, where $\hat{\sigma}_0^2 = \sum_{t=m+1}^{T} \hat{e}_t^2/(T-m)$ is the QMLE of the constant variance based on the OLS residuals $\hat{e}_{m+1}, ..., \hat{e}_T$ of (5). As explained previously, the parameter δ as well as those in $\boldsymbol{b}$ are unidentified when $\boldsymbol{a} = \mathbf{0}$ so their values can be fixed arbitrarily at zero in $\hat{\boldsymbol{\theta}}_0(\boldsymbol{Y})$. The zero values for $\boldsymbol{b}$ are also the correct choice when the null is expressed as $\boldsymbol{a} = \mathbf{0}$, $\boldsymbol{b} = \mathbf{0}$. Note that $\hat{\boldsymbol{\theta}}_0(\boldsymbol{Y})$ is of the same dimension as $\boldsymbol{\theta}$.

The QLR statistic for H_0 against the GARCH alternative H_1 can then be written in the usual way as

$$QLR(\boldsymbol{Y}) = 2\Big(QL\big(\hat{\boldsymbol{\theta}}_1(\boldsymbol{Y}); \boldsymbol{Y}\big) - QL\big(\hat{\boldsymbol{\theta}}_0(\boldsymbol{Y}); \boldsymbol{Y}\big)\Big) \tag{6}$$

where $QL(\hat{\boldsymbol{\theta}}_1(\boldsymbol{Y}); \boldsymbol{Y})$ and $QL(\hat{\boldsymbol{\theta}}_0(\boldsymbol{Y}); \boldsymbol{Y})$ are the quasi-log likelihoods of $\boldsymbol{Y}$ in (4) evaluated at the unrestricted and H_0-restricted QMLEs, respectively.

To set the stage for the general approach developed next, it is useful to introduce a null hypothesis which also specifies values for $\boldsymbol{\beta}$. Consider such a point null hypothesis expressed as

$$H_0(\boldsymbol{\beta}_0) : \boldsymbol{a} = \mathbf{0},\ \boldsymbol{\beta} = \boldsymbol{\beta}_0 \tag{7}$$

where $\boldsymbol{\beta}_0 \in \mathcal{B}$ are the specified values. Here $\mathcal{B}$ denotes a set of admissible values for $\boldsymbol{\beta}$ that are compatible with H_0. When the values in $\boldsymbol{\beta}$ are treated as free parameters, the null hypothesis of conditional homoskedasticity under test is

$$H_0 : \bigcup_{\boldsymbol{\beta}_0 \in \mathcal{B}} H_0(\boldsymbol{\beta}_0) \tag{8}$$

the union of (7) taken over $\mathcal{B}$. The expression in (8) makes clear that the elements of $\boldsymbol{\beta}$ are nuisance parameters in the general setting since they are not constrained to a single value under H_0. The remainder of this subsection develops an exact (quasi-) likelihood-based test of (7). The results established along the way are the building blocks of the general procedure for testing (8), presented in the subsequent subsections.

The construction of MC critical regions for the original sample statistic in (6) makes use of the transformed variables:

$$\varepsilon_t(\boldsymbol{\beta}_0) = y_t - \boldsymbol{x}_t'\boldsymbol{\beta}_0 \tag{9}$$

defined for $t = 1, ..., T$ at the specified values for $\boldsymbol{\beta}_0$ in (7). A reference distribution for the QLR statistic in (6) can be generated through a permutation principle upon noticing that the variables in (9) become $\varepsilon_t(\boldsymbol{\beta}_0) = c + \varepsilon_t$ under $H_0(\boldsymbol{\beta}_0)$. This implies that $\varepsilon_1(\boldsymbol{\beta}_0), ..., \varepsilon_T(\boldsymbol{\beta}_0)$ forms a collection of i.i.d. (i.e. exchangeable) random variables, meaning that for every permutation $d_1, ..., d_T$ of the integers $1, ..., T$,

$$\big(\varepsilon_1(\boldsymbol{\beta}_0), ..., \varepsilon_T(\boldsymbol{\beta}_0)\big) \stackrel{d}{=} \big(\varepsilon_{d_1}(\boldsymbol{\beta}_0), ..., \varepsilon_{d_T}(\boldsymbol{\beta}_0)\big) \tag{10}$$

where the symbol $\stackrel{d}{=}$ stands for the equality in distribution. Essentially, the order of exchangeable observations is irrelevant so that their joint distribution is unaffected when they are permuted. Note that (10) is invariant to the true values of c and σ^2. Under the GARCH alternative, however, the order of $\varepsilon_1(\boldsymbol{\beta}_0), ..., \varepsilon_T(\boldsymbol{\beta}_0)$ does matter owing to serial dependence in the conditional variance process and (10) no longer holds in that case.

Let $\tilde{\varepsilon}_t(\boldsymbol{\beta}_0) = \varepsilon_{d_t}(\boldsymbol{\beta}_0)$ and consider the collection of random variables $\tilde{y}_1(\boldsymbol{\beta}_0), ..., \tilde{y}_T(\boldsymbol{\beta}_0)$, where

$$\tilde{y}_t(\boldsymbol{\beta}_0) = \boldsymbol{x}_t'\boldsymbol{\beta}_0 + \tilde{\varepsilon}_t(\boldsymbol{\beta}_0) \tag{11}$$

for $t = 1, ..., T$. It follows, under $H_0(\boldsymbol{\beta}_0)$ in (7) and given $\boldsymbol{x}_1, ..., \boldsymbol{x}_T$, that

$$\big(y_1, ..., y_T\big) \stackrel{d}{=} \big(\tilde{y}_1(\boldsymbol{\beta}_0), ..., \tilde{y}_T(\boldsymbol{\beta}_0)\big) \tag{12}$$

for each of the $T!$ possible realizations of $\tilde{y}_1(\boldsymbol{\beta}_0), ..., \tilde{y}_T(\boldsymbol{\beta}_0)$. From Theorem 1.3.7 in Randles and Wolfe (1979), it is known that if $\boldsymbol{Z}_1 \stackrel{d}{=} \boldsymbol{Z}_2$ and $U(\cdot)$ is a measurable function (possibly vector-valued) defined on the common support of $\boldsymbol{Z}_1$ and $\boldsymbol{Z}_2$, then $U(\boldsymbol{Z}_1) \stackrel{d}{=} U(\boldsymbol{Z}_2)$. This theorem and the distributional equality in (12) imply that

$$\begin{bmatrix} \hat{\boldsymbol{\theta}}_0(\boldsymbol{Y}) \\ \hat{\boldsymbol{\theta}}_1(\boldsymbol{Y}) \end{bmatrix} \stackrel{d}{=} \begin{bmatrix} \hat{\boldsymbol{\theta}}_0\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big) \\ \hat{\boldsymbol{\theta}}_1\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big) \end{bmatrix} \tag{13}$$

where $\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0) = \big(\tilde{y}_1(\boldsymbol{\beta}_0), ..., \tilde{y}_T(\boldsymbol{\beta}_0)\big)$ denotes a non-parametric "bootstrap" sample generated according to the permutation principle in (11). The same argument yields the following result.

Proposition 1. *Let $\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)$ denote a bootstrap sample generated according to (11) for a given random permutation of the integers $1, ..., T$, and consider the QLR statistic:*

$$QLR\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big) = 2\left\{QL\Big(\hat{\boldsymbol{\theta}}_1\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big); \tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\Big) - QL\Big(\hat{\boldsymbol{\theta}}_0\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big); \tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\Big)\right\}$$

where $QL\Big(\hat{\boldsymbol{\theta}}_1\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big); \tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\Big)$ and $QL\Big(\hat{\boldsymbol{\theta}}_0\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big); \tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\Big)$ are the quasi-log likelihoods of $\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)$ evaluated at the QMLEs $\hat{\boldsymbol{\theta}}_1\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big)$ and $\hat{\boldsymbol{\theta}}_0\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big)$, respectively. Under the point null hypothesis $H_0(\boldsymbol{\beta}_0)$ in (7) and conditional on $\boldsymbol{x}_1, ..., \boldsymbol{x}_T$, the $T!$ values of $QLR\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big)$ that can be obtained from all the possible permutations of the integers $1, ..., T$ are equally likely values for $QLR(\boldsymbol{Y})$.

For the equally likely property in (13) and hence Proposition 1 to hold, it is important that the numerical maximization of $QL\big(\boldsymbol{\theta}; \tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big)$ to get $\hat{\boldsymbol{\theta}}_1\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big)$ be performed in exactly the same manner as was done to get $\hat{\boldsymbol{\theta}}_1(\boldsymbol{Y})$ with the original sample. In particular, this *ceteris paribus* condition means that the only difference is that instead of the original sample $\boldsymbol{Y}$, the bootstrap sample $\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)$ is used to compute and numerically maximize the quasi-log likelihood function in (4). All other aspects of the optimization problem must obviously remain unchanged (e.g. numerical optimization method, choice of initial parameter values, parameter bounds, GARCH initialization scheme, step size, stopping rule, etc.). For instance, if the choice of initial values depends on the data $\boldsymbol{Y}$, then the initial values for the numerical maximization of $QL\big(\boldsymbol{\theta}; \tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big)$ should depend in a completely analogous fashion on $\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)$. Of course, the estimates of the conditional mean parameters that go into $\hat{\boldsymbol{\theta}}_0\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big)$ can be obtained simply via an OLS regression as in (5) and these in turn yield the variance estimate, as before.

The equally likely result established in Proposition 1 forms the basis for a one-sided test of $H_0(\boldsymbol{\beta}_0)$ in (7) using the MC test technique (Barnard, 1963; Birnbaum, 1974; Dwass, 1957). Since large values of the QLR statistic in (6) are more probable under the maintained GARCH alternative, the null hypothesis is rejected if the value of $QLR(\boldsymbol{Y})$ falls in a set C_α containing the $T!\alpha$ largest values of $QLR\big(\tilde{\boldsymbol{Y}}(\boldsymbol{\beta}_0)\big)$ that can be obtained from the class of all permutations. Determination of the critical set C_α by direct counting is obviously impractical. The MC technique avoids this difficulty while still yielding exact p-values. The basic idea is to draw random samples from the permutation distribution and compute the relative frequencies of the resampled QLR statistics.

The MC test proceeds by generating $N-1$ random bootstrap samples $\tilde{\boldsymbol{Y}}_1(\boldsymbol{\beta}_0)$, ..., $\tilde{\boldsymbol{Y}}_{N-1}(\boldsymbol{\beta}_0)$, each one according to (11). With each such sample, the QLR statistic is computed yielding $QLR\big(\tilde{\boldsymbol{Y}}_i(\boldsymbol{\beta}_0)\big)$, for $i = 1, ..., N-1$. Note that the permutation distribution of the QLR statistic is discrete, so ties among the resampled values can occur, at least theoretically. A test with size α can be obtained by applying the following tie-breaking rule (Dufour, 2006). Draw N i.i.d. variates U_i, $i = 1, ..., N$, from a continuous uniform distribution on $[0, 1]$, independently of the QLR statistics, and compute the lexicographic rank of $\big(QLR(\boldsymbol{Y}), U_N\big)$ according to

$$\begin{aligned}\tilde{R}_N(\boldsymbol{\beta}_0) = 1 \quad &+ \quad \sum_{i=1}^{N-1} \mathbb{I}\left[QLR(\boldsymbol{Y}) > QLR\big(\tilde{\boldsymbol{Y}}_i(\boldsymbol{\beta}_0)\big)\right] \\ &+ \quad \sum_{i=1}^{N-1} \mathbb{I}\left[QLR(\boldsymbol{Y}) = QLR\big(\tilde{\boldsymbol{Y}}_i(\boldsymbol{\beta}_0)\big)\right] \times \mathbb{I}\big[U_N > U_i\big] \end{aligned} \tag{14}$$

where $\mathbb{I}[A]$ is the indicator function of event A.

Now observe that the pairs:

$$\left(QLR\big(\tilde{\boldsymbol{Y}}_1(\boldsymbol{\beta}_0)\big), U_1\right), ..., \left(QLR\big(\tilde{\boldsymbol{Y}}_{N-1}(\boldsymbol{\beta}_0)\big), U_{N-1}\right), \big(QLR(\boldsymbol{Y}), U_N\big)$$

are exchangeable under $H_0(\boldsymbol{\beta}_0)$, meaning that the lexicographic ranks are uniformly distributed over the integers $1, ..., N$. So the MC p-value can be defined as

$$\tilde{p}_N(\boldsymbol{\beta}_0) = \frac{N - \tilde{R}_N(\boldsymbol{\beta}_0) + 1}{N} \tag{15}$$

where $\tilde{R}_N(\boldsymbol{\beta}_0)$ is the lexicographic rank of $\big(QLR(\boldsymbol{Y}), U_N\big)$ in (14). If αN is an integer, then the critical region $\tilde{p}_N(\boldsymbol{\beta}_0) \le \alpha$ has exactly size α in the sense that

$$\Pr\left[\tilde{p}_N(\boldsymbol{\beta}_0) \le \alpha \mid \boldsymbol{x}_1, ..., \boldsymbol{x}_T\right] = \alpha$$

under the point null hypothesis $H_0(\boldsymbol{\beta}_0)$ in (7).

2.1.1 Maximized MC test

In order to test a null hypothesis like the one in (8) which contains several distributions, one can appeal to a *minimax* argument stated as: "reject the null hypothesis whenever for all admissible values of the nuisance parameters under the null, the corresponding point null hypothesis is rejected;" see Savin (1984). In practice, this approach consists of maximizing the p-value of the original sample test statistic $QLR(\boldsymbol{Y})$ over the set of nuisance parameters. The rationale is that $\sup_{\boldsymbol{\beta}_0 \in \mathcal{B}} \left[\tilde{p}_N(\boldsymbol{\beta}_0)\right] \leq \alpha \Rightarrow \tilde{p}_N(\boldsymbol{\beta}) \leq \alpha$, where the latter is the MC p-value of $QLR(\boldsymbol{Y})$ based on the true parameter values. Moreover, $\Pr\left[\tilde{p}_N(\boldsymbol{\beta}_0) \leq \alpha \mid \boldsymbol{x}_1, ..., \boldsymbol{x}_T\right] = \alpha$, under $H_0(\boldsymbol{\beta}_0)$ and for each $\boldsymbol{\beta}_0 \in \mathcal{B}$. So if αN is an integer, then

$$\Pr\left[\sup_{\boldsymbol{\beta}_0 \in \mathcal{B}} \left[\tilde{p}_N(\boldsymbol{\beta}_0)\right] \leq \alpha \,\middle|\, \boldsymbol{x}_1, ..., \boldsymbol{x}_T\right] \leq \alpha$$

under H_0, meaning that the test of (8) which rejects whenever $\sup_{\boldsymbol{\beta}_0 \in \mathcal{B}} \left[\tilde{p}_N(\boldsymbol{\beta}_0)\right] \leq \alpha$ is guaranteed to have level α. The computation of the MMC p-value can be laborious, so it will be used only as a last resort if the simpler local MC and bounds MC tests presented next are inconclusive.

2.1.2 Local MC test

The local MC (LMC) test focuses on the point null hypothesis in (7) evaluated at the QMLE obtained by OLS under conditional homoskedasticity, written here as

$$H_0(\hat{\boldsymbol{\beta}}_0) : \boldsymbol{a} = \boldsymbol{0},\ \boldsymbol{\beta} = \hat{\boldsymbol{\beta}}_0 \tag{16}$$

where $\hat{\boldsymbol{\beta}}_0$, by construction, is compatible with H_0. The associated p-value computed according to (15) is $\tilde{p}_N(\hat{\boldsymbol{\beta}}_0)$ and the LMC test consists of rejecting $H_0(\hat{\boldsymbol{\beta}}_0)$ at level α if $\tilde{p}_N(\hat{\boldsymbol{\beta}}_0) \leq \alpha$. In light of (8) seen as a union of point null hypotheses, it is clear that a rejection of (16) by the LMC test does not imply that H_0 is false. On the flip side, if the LMC test is not significant, then the MMC test would not be significant either and the decision is to accept the null hypothesis of conditional homoskedasticity. The logic of this decision not to reject H_0 when $\tilde{p}_N(\hat{\boldsymbol{\beta}}_0) > \alpha$ follows from the fact that $H_0(\hat{\boldsymbol{\beta}}_0) \subseteq H_0$.

2.1.3 Bounds MC test

If the LMC test indicates a rejection, it is possible through a bounds MC (BMC) test to confirm that H_0 should indeed be rejected without any need for the MMC test. Just like the LMC test, the BMC approach also focuses on (16) but introduces another statistic whose distribution bounds that of $QLR(\boldsymbol{Y})$.

When $H_0(\hat{\boldsymbol{\beta}}_0)$ in (16) is imposed, the only free parameters are c and σ^2 which can be estimated as $\hat{c}_0^* = \sum_{t=m+1}^{T} \varepsilon_t(\hat{\boldsymbol{\beta}}_0)/(T-m)$ and $\hat{\sigma}_0^{*2} = \sum_{t=m+1}^{T} \big(\varepsilon_t(\hat{\boldsymbol{\beta}}_0) - \hat{c}_0^*\big)^2/(T-m-1)$, where $\varepsilon_t(\hat{\boldsymbol{\beta}}_0) = y_t - \boldsymbol{x}_t'\hat{\boldsymbol{\beta}}_0$ is the transformation in (9) evaluated at the point $\hat{\boldsymbol{\beta}}_0$. The collection of these $H_0(\hat{\boldsymbol{\beta}}_0)$-restricted parameter estimates are stacked into $\hat{\boldsymbol{\theta}}_0^*(\boldsymbol{Y}) = (\hat{c}_0^*, 0, \hat{\boldsymbol{\beta}}_0', \hat{\sigma}_0^{*2}, \mathbf{0}', \mathbf{0}')'$ and the QLR statistic for $H_0(\hat{\boldsymbol{\beta}}_0)$ against the GARCH alternative is defined as

$$QLR^*(\boldsymbol{Y}) = 2\Big(QL\big(\hat{\boldsymbol{\theta}}_1(\boldsymbol{Y}); \boldsymbol{Y}\big) - QL\big(\hat{\boldsymbol{\theta}}_0^*(\boldsymbol{Y}); \boldsymbol{Y}\big)\Big) \tag{17}$$

The fact that $\hat{\boldsymbol{\theta}}_0^*(\boldsymbol{Y})$ is a more restricted estimate of $\boldsymbol{\theta}$ than $\hat{\boldsymbol{\theta}}_0(\boldsymbol{Y})$ implies that $QL\big(\hat{\boldsymbol{\theta}}_0^*(\boldsymbol{Y}); \boldsymbol{Y}\big) \leq QL\big(\hat{\boldsymbol{\theta}}_0(\boldsymbol{Y}); \boldsymbol{Y}\big) \leq QL\big(\hat{\boldsymbol{\theta}}_1(\boldsymbol{Y}); \boldsymbol{Y}\big)$, since a restricted maximum cannot be greater than a less restricted one. In turn, this means that

$$0 \leq \Big(QL\big(\hat{\boldsymbol{\theta}}_1(\boldsymbol{Y}); \boldsymbol{Y}\big) - QL\big(\hat{\boldsymbol{\theta}}_0(\boldsymbol{Y}); \boldsymbol{Y}\big)\Big) \leq \Big(QL\big(\hat{\boldsymbol{\theta}}_1(\boldsymbol{Y}); \boldsymbol{Y}\big) - QL\big(\hat{\boldsymbol{\theta}}_0^*(\boldsymbol{Y}); \boldsymbol{Y}\big)\Big) \tag{18}$$

which leads to the following result.

Proposition 2. *Given the model in (1) and (2) with the GARCH specification in (3), consider the QLR statistics $QLR(\boldsymbol{Y})$ in (6) and $QLR^*(\boldsymbol{Y})$ in (17). The inequality in (18) implies that*

$$\Pr\big[QLR(\boldsymbol{Y}) > \zeta\big] \leq \Pr\big[QLR^*(\boldsymbol{Y}) > \zeta\big]$$

for any $\zeta \in \mathbb{R}$.

To see the usefulness of this result, let ζ_α and ζ_α^* be critical values defined via $\Pr[QLR(\boldsymbol{Y}) > \zeta_\alpha] = \alpha$ under H_0, and $\Pr[QLR^*(\boldsymbol{Y}) > \zeta_\alpha^*] = \alpha$ under $H_0(\hat{\boldsymbol{\beta}}_0)$. Proposition 2 implies that $\zeta_\alpha \le \zeta_\alpha^*$, which in turn means that $\Pr[QLR(\boldsymbol{Y}) > \zeta_\alpha^*] \le \alpha$ when H_0 is true. Proposition 2 further implies that if $QLR(\boldsymbol{Y}) > \zeta_\alpha^*$ then $QLR(\boldsymbol{Y}) > \zeta_\alpha$; i.e., if the QLR bounds test based on ζ_α^* is significant, then for sure the exact QLR test involving the unknown ζ_α is also significant at level α.

The computation of the BMC p-value now proceeds very much as before. Specifically, it begins by generating $N-1$ i.i.d. bootstrap samples $\tilde{\boldsymbol{Y}}_1(\hat{\boldsymbol{\beta}}_0), ..., \tilde{\boldsymbol{Y}}_{N-1}(\hat{\boldsymbol{\beta}}_0)$, each one according to (11) evaluated at $\hat{\boldsymbol{\beta}}_0$. Each of those samples serves to compute $QLR^*\big(\tilde{\boldsymbol{Y}}_i(\hat{\boldsymbol{\beta}}_0)\big)$ for $i = 1, ..., N-1$. The tie-breaking rule is then applied by drawing i.i.d. uniform variates over the interval $[0, 1]$, say U_i, for $i = 1, ..., N$, independently of all the QLRs. The lexicographic rank of $\big(QLR(\boldsymbol{Y}), U_N\big)$ among the bootstrap pairs is then computed according to

$$\begin{aligned}\tilde{R}_N^*(\hat{\boldsymbol{\beta}}_0) = 1 \quad &+ \quad \sum_{i=1}^{N-1} \mathbb{I}\left[QLR(\boldsymbol{Y}) > QLR^*\big(\tilde{\boldsymbol{Y}}_i(\hat{\boldsymbol{\beta}}_0)\big)\right] \\ &+ \quad \sum_{i=1}^{N-1} \mathbb{I}\left[QLR(\boldsymbol{Y}) = QLR^*\big(\tilde{\boldsymbol{Y}}_i(\hat{\boldsymbol{\beta}}_0)\big)\right] \times \mathbb{I}\big[U_N > U_i\big] \qquad (19)\end{aligned}$$

It is important to note that this ranking is based on a comparison of the original QLR statistic $QLR(\boldsymbol{Y})$ in (6) with the bounding statistics $QLR^*\big(\tilde{\boldsymbol{Y}}_i(\hat{\boldsymbol{\beta}}_0)\big)$ computed according to (17). The BMC p-value can be defined as

$$\tilde{p}_N^*(\hat{\boldsymbol{\beta}}_0) = \frac{N - \tilde{R}_N^*(\hat{\boldsymbol{\beta}}_0) + 1}{N}$$

where $\tilde{R}_N^*(\hat{\boldsymbol{\beta}}_0)$ is the lexicographic rank of $\big(QLR(\boldsymbol{Y}), U_N\big)$ in (19). If αN is an integer, then the critical region $\tilde{p}_N^*(\hat{\boldsymbol{\beta}}_0) \le \alpha$ has exactly level α in the sense that

$$\Pr\left[\tilde{p}_N^*(\hat{\boldsymbol{\beta}}_0) \le \alpha \mid \boldsymbol{x}_1, ..., \boldsymbol{x}_T\right] \le \alpha$$

when the null hypothesis H_0 in (8) is true.

The BMC test therefore consists of rejecting the null hypothesis of conditional homoskedasticity H_0 at level α if $\tilde{p}_N^*(\hat{\boldsymbol{\beta}}_0) \le \alpha$. This follows from Proposition 2 which shows that the (unknown)

p-value of $QLR(\boldsymbol{Y})$ under H_0 is bounded from above by its computable p-value under the $H_0(\hat{\boldsymbol{\beta}}_0)$-distribution of $QLR^*(\boldsymbol{Y})$; i.e., if the BMC p-value indicates a rejection, then for sure the p-value of $QLR(\boldsymbol{Y})$ under its unknown distribution would also indicate a rejection of H_0 in (8). A non-rejection by the BMC test (i.e. $p_N^*(\hat{\boldsymbol{\beta}}_0) > \alpha$) on the other hand does not necessarily mean that H_0 is true. Fortunately, the MMC, LMC, and BMC tests complement one another to form a comprehensive testing strategy, summarized next.

2.1.4 Summary of GARCH test procedure

Consider the model specification in (1) and (2) with error terms described by the benchmark GARCH model in (3). If values for $\boldsymbol{\beta}$ in (1) are specified (e.g. $\boldsymbol{\beta} = \mathbf{0}$), then the relevant null hypothesis of conditional homoskedasticity is $H_0(\boldsymbol{\beta}_0)$ in (7). In this case, obtain the size-α MC p-value in (15) and reject $H_0(\boldsymbol{\beta}_0)$ if: MC p-value $\leq \alpha$. Otherwise accept $H_0(\boldsymbol{\beta}_0)$.

If the values of $\boldsymbol{\beta}$ in (1) are unknown, then the null hypothesis of constant conditional variances is given by H_0 in (8) and the test procedure proceeds according to the following steps:

1. Obtain the LMC p-value of $QLR(\boldsymbol{Y})$ and accept H_0 if: LMC p-value $> \alpha$. Otherwise proceed to Step 2.

2. Obtain the BMC p-value of $QLR(\boldsymbol{Y})$ and reject H_0 if: BMC p-value $\leq \alpha$. Otherwise proceed to Step 3.

3. If the LMC p-value $\leq \alpha <$ BMC p-value, obtain the MMC p-value of $QLR(\boldsymbol{Y})$ and reject H_0 if: MMC p-value $\leq \alpha$. Otherwise accept H_0.

This procedure guarantees that $\Pr[\, H_0 \text{ rejected} \mid \boldsymbol{x}_1, ..., \boldsymbol{x}_T] \leq \alpha$ when H_0 is true, since an LMC test rejection in Step 1 does not affect the exactness of the BMC test in Step 2, and, moreover, a rejection in Step 1 and a non-rejection in Step 2 does not affect the exactness of the MMC test in Step 3.

Before closing this section, it is important to remark that the same set of random permutations of the integers $1, ..., T$ used to obtain the bootstrap samples and the same set of uniform draws used in the tie-breaking rule should be reused when computing the LMC, BMC, and MMC p-values in order to avoid conflicting answers. This means that the MMC p-value is found by varying the candidate points $\boldsymbol{\beta}_0$ only; i.e., the randomness in $\tilde{p}_N(\boldsymbol{\beta}_0)$ should be held constant across candidate points. A simple way to ensure this is to reset the seed of the random number generator to the same value before Step 1, before Step 2, and before each considered $\boldsymbol{\beta}_0$ in Step 3.

3 Simulation evidence

This section presents the results of simulation experiments to compare the performance of the new GARCH test procedure with several standard tests. The MC procedure is implemented with $N = 20$, which is sufficient for a test at the nominal 5% level. The benchmarks for comparison purposes include the commonly used LM tests of Engle (1982), Lee and King (1993), and Demos and Sentana (1998). The experiments examine two prominent model specifications: (i) the symmetric GARCH model of Bollerslev (1986); and (ii) the asymmetric GARCH model of Glosten, Jagannathan, and Runkle (1993) (GJR).

The conditional mean specification first takes the basic GARCH-in-mean form:

$$y_t = \mu + \delta \log(\sigma_t^2) + \varepsilon_t \tag{20}$$

which does not involve any nuisance parameters, and then it assumes a form that also includes an exogenous covariate:

$$y_t = \mu + \delta \log(\sigma_t^2) + \beta x_t + \varepsilon_t \tag{21}$$

where β is treated as a free parameter along with μ and δ. The error terms in (20) and (21) are given by $\varepsilon_t = \sigma_t \eta_t$, for $t = 1, ..., T$, with innovations η_t drawn either from the $N(0, 1)$ distribution or from the $t(6)$ distribution (standardized to have unit variance). The exogenous regressor x_t in (21) is generated as i.i.d. according to the standard normal distribution and $\beta = 1$. This choice is rather innocuous since the test procedure is conditional on the x_ts. In both (20) and (21) the intercept is set as $\mu = 0$; the risk premium parameter takes values $\delta = 0$, 0.5, 1.0; and sample sizes $T = 120$, 240 are considered. These sample sizes correspond to the 10 and 20 years of monthly returns used with both of the above specifications for the empirical application in Section 4.

The first model for the conditional variance of the error terms in (20) and (21) is the GARCH(1,1) version of (3). For convenience, it is given again here as

$$\sigma_t^2 = \omega + a_1 \varepsilon_{t-1}^2 + b_1 \sigma_{t-1}^2 \tag{22}$$

where $\omega = 0.1$; $a_1 = 0$, 0.05, 0.10; and $b_1 = 0.80$, 0.85, 0.89. This choice of parameter values is motivated by the estimates typically reported in the empirical literature. Indeed, the estimate of a_1 is usually found to be small, while the estimate of b_1 is much larger with their sum $a_1 + b_1$ between 0.9 and 1 (cf. Zivot, 2009). The first variance term is set as $\sigma_1^2 = \omega/(1 - a_1 - b_1)$ so that the null hypothesis of conditional homoskedasticity is represented by $a_1 = 0$ in (22). The second model examined is the GJR-GARCH(1,1) specification:

$$\sigma_t^2 = \omega + a_1 \varepsilon_{t-1}^2 + \gamma_1 \varepsilon_{t-1}^2 \mathbb{I}[\varepsilon_{t-1} < 0] + b_1 \sigma_{t-1}^2 \tag{23}$$

where the parameters are varied as $\omega = 0.1$; $a_1 = 0$, 0.05, 0.10; $\gamma_1 = 0$, 0.04, 0.08; and $b_1 = 0.80$, 0.85. With the GJR specification, the first variance term is set as $\sigma_1^2 = \omega/(1 - a_1 - \gamma_1/2 - b_1)$ and the null hypothesis of constant conditional variances occurs when $a_1 = \gamma_1 = 0$ in (23).

The benchmark LM tests of conditional homoskedasticity used for comparisons are each based on the OLS residuals $\hat{\varepsilon}_t$, $t = 1, ..., T$, from a first-stage regression of y_t on a constant in the case of (20); and on a constant and x_t when (21) is maintained. The usual Engle (1982) (E) test for ARCH(1) effects is computed as T times R^2, the coefficient of determination in a second-stage regression of $\hat{\varepsilon}_t^2$ on a constant and $\hat{\varepsilon}_{t-1}^2$. Under the null hypothesis of conditional homoskedasticity, this E statistic follows an asymptotic $\chi^2(1)$ distribution. The results in Lee (1991) show that the same test is appropriate against GARCH(1,1) alternatives. An equivalent version of the E test can also be obtained from the squared t-statistic in the second-stage regression (Demos and Sentana, 1998).

Lee and King (1993) suggest another LM test which exploits the one-sided nature of GARCH alternatives. Their test rejects the null hypothesis of conditional homoskedasticity for large values of

$$LK = \frac{\left\{(T-1)\sum_{t=2}^{T}\left[(\hat{\varepsilon}_t/\hat{\sigma})^2-1\right]\hat{\varepsilon}_{t-1}^2\right\} \Big/ \left\{\sum_{t=2}^{T}\left[(\hat{\varepsilon}_t/\hat{\sigma})^2-1\right]^2\right\}^{1/2}}{\left\{(T-1)\sum_{t=2}^{T}\hat{\varepsilon}_{t-1}^4-\left(\sum_{t=2}^{T}\hat{\varepsilon}_{t-1}^2\right)^2\right\}^{1/2}} \tag{24}$$

where $\hat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\hat{\varepsilon}_t^2$. The Lee-King statistic in (24) is asymptotically distributed as a $N(0,1)$ variate under the null hypothesis of no GARCH effects.

Demos and Sentana (1998) (DS) also propose a one-sided LM test for GARCH effects based on a Kuhn-Tucker approach. As with the Engle (1982) and Lee and King (1993) tests, the DS approach for GARCH(1,1) alternatives is also based on the auxiliary regression of the squared residuals on a constant and the lagged squared residuals. The DS test statistic is then the squared t-statistic if the second-stage OLS regression parameter estimate is positive, or zero otherwise. The asymptotic distribution of this statistic is given by an equiprobable mixture of $\chi^2(0)$ and $\chi^2(1)$. Since $\chi^2(0)$ is a random variable which equals zero with probability 1, the p-value associated with a computed DS statistic value is simply given by $0.5\Pr[\chi^2(1) > DS]$ in the GARCH(1,1) case. As Demos and Sentana explain, the one-sided LK and DS tests should be more powerful than the two-sided E test because the latter ignores that the ARCH parameters are non-negative under the alternative hypothesis. It is important to realize that the MC test procedure proposed here—based on a directional QLR statistic—is also one-sided in nature since the GARCH alternative is explicitly taken into account.

Tables 1–3 show the empirical rejection probabilities, in percentages, of the MC test procedure as well as those of the E, LK, and DS tests. The MMC step of the test procedure was performed by grid search over a uniform interval such that β_0 took values in $[\hat{\beta}_0-3, \hat{\beta}_0+3]$, where $\hat{\beta}_0$ is the QMLE under H_0 obtained via the OLS regression in (5). (The results did not change beyond that interval.)

The reported rejection rates are based on 1000 replications of each data-generating process (DGP) configuration. Table 1 reports the test results for the conditional mean specification in (20), and Tables 2 and 3 for the one in (21). Specifically, Table 1 shows the size and power of the MC test procedure with $T = 120$ assuming in turn the symmetric GARCH and then the asymmetric GJR-GARCH model, and when the true innovations come from either the $N(0,1)$ or $t(6)$ distribution. Table 2 focuses on the symmetric GARCH model with $T = 120$, 240, and Table 3 does the same but for the GJR-GARCH model. In Tables 1–3, the null hypothesis of conditional homoskedasticity is represented by $a_1 = 0$ (or $a_1 = \gamma_1 = 0$) on the first line of each panel. The following lines where $a_1 \neq 0$ reveal the power of the tests to detect GARCH effects; the entries set in bold correspond to the most powerful test under each DGP configuration.

The tables show that the MC test respects the nominal 5% level constraint, as expected from the developed theory. In the case of Table 1 when there are no nuisance parameters, the empirical rejection probabilities are close to 5%, while in Tables 2 and 3 the MC test appears conservative with empirical sizes less than 5% owing to the presence of a nuisance parameter. The first lines of each panel show that the E, LK, and DS tests also respect the nominal level constraint quite well. Looking across the three tables one can see that the one-sided LK and DS tests have practically identical power for every considered DGP configuration. Demos and Sentana (1998) also note this finding in the GARCH(1,1) case. The power of the two-sided E test is also seen to be uniformly dominated by one of the three other one-sided tests in all instances.

For alternatives close to the null and when δ is close to zero, the LK and DS tests tend to have better power than the MC test. However, as the ARCH parameter a_1 and the risk premium parameter δ assume values farther away from zero, the MC test has better power than the E-LK-DS group. In fact, the power of the latter tests barely responds to changes in δ. On the contrary, the power of the MC test increases quite dramatically with δ. This is not surprising since the MC test is based on a model that explicitly captures GARCH-in-mean effects.

The simulation evidence also reveals that the power of the MC test increases at a faster rate as the alternative moves away from the null hypothesis. In Panel A of Table 1 when $\delta = 0$ for example, the power of the LK and DS tests doubles from about 14% to nearly 28% as (a_1, b_1) changes from (0.05, 0.89) to (0.10, 0.89), while the power of the MC test more than triples from about 14% to 46%. Similar power increases for the MC test are also seen in Tables 2 and 3, and of course they become even more dramatic for higher values of δ. The tables further show how the power of each test responds to changes in the GARCH parameter b_1. It is interesting to observe that the power of the E-LK-DS group is not very responsive to increases in b_1. For instance, in Panel A of Table 1 when $\delta = 0$, the power of the MC test increases from 28% to 34% to 46% as (a_1, b_1) changes from (0.10, 0.80) to (0.10, 0.85) to (0.10, 0.89), while the power of the E-LK-DS group of LM tests stays about the same in the 18–27% range. Again, this occurs because the MC test is based on an explicit GARCH specification, while the LM tests detect ARCH-like effects.

Tables 1–3 show that the power of all four tests is generally lower with $t(6)$ innovations than with normal ones. This finding in entirely consistent with the results of Bollerslev and Wooldridge (1992), Lee and King (1993), and Demos and Sentana (1998), and simply reflects the fact that the "optimality" of GARCH tests derived from a Gaussian likelihood gets lost when moving to a quasi-maximum likelihood context. Comparing Tables 2 and 3 against Table 1 reveals that the MC test suffers a relative power loss in the presence of nuisance parameters. It is also clear that all the GARCH tests benefit from an increased sample size at a fixed alternative, as expected. A close examination of Table 3 shows that the power of all four tests increases with the GJR asymmetry parameter γ_1. As before though, the power of the E-LK-DS group is not very responsive to stronger GARCH-in-mean effects and the power of the MC test is seen again to increase at a faster rate and surpass the LM tests as the alternative moves farther away from the null hypothesis point $a_1 = \gamma_1 = 0$.

The simulation results in Tables 1–3 are all the more remarkable considering that the MC procedure used only $N - 1 = 19$ bootstrap replications. So when compared to the omnibus E, LK, and DS tests, the evidence presented here clearly shows that with very little extra computational effort a potentially far more powerful GARCH test can be obtained.

4 Application to Fama-French returns

The new MC test procedure is illustrated by an application to the well-known Fama and French (1993) portfolio data, which are routinely used in empirical asset pricing studies. Specifically, the data consist of monthly returns on 10 portfolios formed on size for the 10-year period covering January 2001 to December 2010 ($T = 120$) and the 20-year period from January 1991 to December 2010 ($T = 240$). The models examined here are the ones in (20) and (21), where y_t denotes in turn one of the 10 portfolio excess returns and x_t is the excess return of a value-weighted stock market index of all stocks listed on the NYSE, AMEX, and NASDAQ—the usual market risk factor. These specifications are examined allowing for a GARCH-in-mean effect (δ free) and also forcing δ to equal zero. Here excess returns are those in excess of the rate of return on the one-month U.S. Treasury bill. This model setup is in the spirit of the CAPM of Sharpe (1964) and Lintner (1965), and the data are available at Ken French's online data library.

Table 4 reports the p-values of the MC, E, LK, and DS tests using the portfolio returns over the period from January 2001 to December 2010 in Panels A and C, and from January 1991 to December 2010 in Panels B and D. The results are organized from the smallest decile portfolio (Lo 10) to the highest one (Hi 10). The columns set in bold show important disagreements between the E-LK-DS group and the MC tests, which are performed assuming GARCH(1,1) and GJR-GARCH(1,1) models for the error terms. The MC test procedure was implemented exactly as described in the previous section except that the number of bootstrap replications was increased to $N - 1 = 99$, which means that the smallest possible MC p-value is now 0.01.

The overall picture that emerges from Table 4 is that there is very strong evidence suggesting the presence of GARCH effects in the Fama-French portfolio returns, even when conditioning on the market risk factor. The results for the basic specification shown in Panels A and B consistently reject conditional homoskedasticity. A disagreement between the E-LK-DS group and the MC tests occurs in Decile 2, where the MC tests indicate a rejection of the null hypothesis but not the LM tests. When the market risk factor is conditioned upon, the results in Panel C with $T = 120$ suggest no GARCH effects in Deciles 6–8. Panel D, however, shows that when the sample size is increased to $T = 240$ the evidence of GARCH effects becomes much more compelling on the basis of the MC tests. Indeed, it is only for Decile 8 that the null hypothesis cannot be rejected at conventional levels of significance. The E-LK-DS group agrees with the MC test findings except for Deciles 6, 7, and 9. These findings are interesting because it is often believed that the market risk factor subsumes any return heterogeneity; i.e., that once the market risk factor is conditioned upon, then asset returns should be identically distributed. This is clearly not the case here. In general, the inference results based on the MC test procedure are virtually identical whether or not δ is set to zero. A notable exception occurs for Decile 7 in Panel D, where the MC GARCH test shows a non-rejection (p-value of 0.17) when $\delta = 0$, but then decisively rejects conditional homoskedasticity once the GARCH-in-mean effect is allowed for.

In light of the simulation results in the previous section, a very natural explanation for the disagreements in Table 4 is the lack of power of the LM tests against certain alternatives relative to the QLR-based test procedure. It is interesting to note that the new results presented here stand in sharp contrast to those of Kan and Zhou (2006) who argue that the Fama-French returns have no GARCH effects, but only on the basis of Engle's two-sided LM test. Observe that the results in Table 4 are for each portfolio taken one at a time. Extending the proposed test procedure to a multivariate GARCH context is the subject of ongoing research.

5 Conclusion

This paper has developed a distribution-free Monte Carlo procedure to test the null hypothesis of conditional homoskedasticity in the context of specified GARCH models. The test procedure is guaranteed to control the probability of committing a Type I error in finite samples, despite parameter identification problems that occur under the null. The approach is based on quasi-likelihood ratios and does not make any parametric distributional assumptions about the model's innovation terms. This is very attractive and quite natural since quasi-maximum likelihood estimation is often the method of choice for GARCH models when one does not want to commit to any particular distribution. Luger (2012) describes how to conduct finite-sample bootstrap inference in GARCH models when the specification assumes a parametric distribution.

The feasibility of the new test procedure was illustrated by means of simulation experiments and an empirical application. The results show that the proposed test displays very desirable power properties in comparison to the standard two-sided Lagrange multiplier test and one-sided variants. In particular, the power of the permutation-based QLR test increases at a faster rate and surpasses the LM tests when the alternative moves farther away from the null hypothesis into parameter regions where GARCH effects matter more. The power of the new QLR test also increases dramatically with the GARCH-in-mean parameter, while the LM tests barely respond to changes in that parameter value. These results are in line with what one would expect since the LM tests are omnibus in nature in the sense that they detect departures from the null hypothesis, but not in any specific GARCH direction. On the other hand, the directional QLR approach advocated here is based on the user's explicit model of the GARCH alternative hypothesis.

References

Andrews, D., 2001. Testing when a parameter is on the boundary of the maintained hypothesis. Econometrica 69, 683–734.

Barnard, G., 1963. Comment on 'The spectral analysis of point processes' by M.S. Bartlett. Journal of the Royal Statistical Society (Series B) 25, 294.

Beg, R., Silvapulle, M., Silvapulle, P., 2001. Tests against inequality constraints when some nuisance parameters are present only under the alternative: test of ARCH in ARCH-M models. Journal of Business and Economic Statistics 19, 245–253.

Berkes, I., Horváth, L., Kokoszka, P., 2003. GARCH processes: structure and estimation. Bernoulli 9, 201–227.

Birnbaum, Z., 1974. Computers and unconventional test-statistics. In: Proschan, F., Serfling, R. (Eds.), Reliability and Biometry. SIAM, Philadelphia, pp. 441–458.

Bollerslev, T., 1986. Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31, 307–327.

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

Bollerslev, T., Wooldridge, J., 1992. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. Econometric Reviews 11, 143–172.

Bougerol, P., Picard, N., 1992. Stationarity of GARCH processes and some nonnegative time series. Journal of Econometrics 52, 115–127.

Demos, A., Sentana, E., 1998. Testing for GARCH effects: a one-sided approach. Journal of Econometrics 86, 97–127.

Dufour, J.-M., 2006. Monte Carlo tests with nuisance parameters: a general approach to finite-sample inference and nonstandard asymptotics in econometrics. Journal of Econometrics 133, 443–477.

Dufour, J.-M., Khalaf, L., 2002. Simulation based finite and large sample tests in multivariate regressions. Journal of Econometrics 111, 303–322.

Dufour, J.-M., Khalaf, L., 2003. Finite-sample simulation-based tests in seemingly unrelated regressions. In: Giles, D. (Ed.), Computer-Aided Econometrics. Marcel Dekker, New York, pp. 11–35.

Dufour, J.-M., Khalaf, L., Bernard, J.-T., Genest, I., 2004. Simulation-based finite-sample tests for heteroskedasticity and ARCH effects. Journal of Econometrics 122, 317–347.

Dwass, M., 1957. Modified randomization tests for nonparametric hypotheses. Annals of Mathematical Statistics 28, 181–187.

Engle, R., 1982. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. Econometrica 50, 987–1007.

Engle, R., Lilien, D., Robins, R., 1987. Estimating time varying risk premia in the term structure: the ARCH-M model. Econometrica 55, 391–407.

Fama, E., French, K., 1993. Common risk factors in the returns on stocks and bonds. Journal of Financial Economics 33, 3–56.

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

Francq, C., Zakoïan, J.-M., 2007. Quasi-maximum likelihood estimation in GARCH processes when some coefficients are equal to zero. Stochastic Processes and their Applications 117, 1265–1284.

Francq, C., Zakoïan, J.-M., 2010. GARCH Models: Structure, Statistical Inference and Financial Applications. John Wiley & Sons Ltd.

French, K., Schwert, G., Stambaugh, R., 1987. Expected stock returns and volatility. Journal of Financial Economics 19, 3–29.

Glosten, L., Jagannathan, R., Runkle, D., 1993. On the relation between the expected value and the volatility of the nominal excess return on stocks. Journal of Finance 48, 1779–1801.

Gouriéroux, C., Monfort, A., Trognon, A., 1984. Pseudo maximum likelihood methods: theory. Econometrica 52, 681–700.

Hayashi, F., 2000. Econometrics. Princeton University Press, Princeton, New Jersey.

Hentschel, L., 1995. All in the family: nesting symmetric and asymmetric GARCH models. Journal of Financial Economics 39, 71–104.

Jensen, S., Rahbek, A., 2004a. Asymptotic inference for nonstationary GARCH. Econometric Theory 20, 1203–1226.

Jensen, S., Rahbek, A., 2004b. Asymptotic normality of the QMLE estimator of ARCH in the nonstationary case. Econometrica 72, 641–646.

Kan, R., Zhou, G., 2006. Modeling non-normality using multivariate t: implications for asset pricing. Working Paper, Washington University in St. Louis.

Khalaf, L., Saphores, J.-D., Bilodeau, J.-F., 2003. Simulation-based exact jump tests in models with conditional heteroskedasticity. Journal of Economic Dynamics and Control 28, 531–553.

Lee, J., 1991. A Lagrange multiplier test for GARCH models. Economics Letters 37, 265–271.

Lee, J., King, M., 1993. A locally most mean powerful based score test for ARCH and GARCH regression disturbances. Journal of Business and Economic Statistics 11, 17–27.

Lee, S.-W., Hansen, B., 1994. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. Econometric Theory 10, 29–53.

Lintner, J., 1965. The valuation of risk assets and the selection of risky investments in stock portfolios and capital budgets. Review of Economics and Statistics 47, 13–37.

Luger, R., 2006. Exact permutation tests for non-nested non-linear regression models. Journal of Econometrics 133, 513–529.

Luger, R., 2012. Finite-sample bootstrap inference in GARCH models with heavy-tailed innovations. Computational Statistics and Data Analysis 56, 3198–3211.

Lumsdaine, R., 1996. Consistency and asymptotic normality of the quasi-maximum likelihood estimator in IGARCH(1,1) and covariance stationary GARCH(1,1) models. Econometrica 64, 575–596.

Nelson, D., 1990. Stationarity and persistence in the GARCH(1,1) model. Econometric Theory 6, 318–334.

Nelson, D., 1991. Conditional heteroskedasticity in asset returns: a new approach. Econometrica 59, 347–370.

Nelson, D., Cao, C., 1992. Inequality constraints in the univariate GARCH model. Journal of Business and Economic Statistics 10, 229–235.

Politis, D., 2004. A heavy-tailed distribution for ARCH residuals with application to volatility prediction. Annals of Economics and Finance 5, 283–298.

Randles, R., Wolfe, D., 1979. Introduction to the Theory of Nonparametric Statistics. Wiley, New York.

Savin, N., 1984. Multiple hypothesis testing. In: Griliches, Z., Intriligator, M. (Eds.), Handbook of Econometrics. North-Holland, Amsterdam, pp. 827–879.

Sentana, E., 1995. Quadratic ARCH models. Review of Economic Studies 62, 639–661.

Sharpe, W., 1964. Capital asset prices: a theory of market equilibrium under conditions of risk. Journal of Finance 19, 425–442.

Weiss, A., 1986. Asymptotic theory for ARCH models: estimation and testing. Econometric Theory 2, 107–131.

Zivot, E., 2009. Practical issues in the analysis of univariate GARCH models. In: Andersen, T., Davis, R., Kreiss, J.-P., Mikosch, T. (Eds.), Handbook of Financial Time Series. Springer-Verlag, Berlin, pp. 113–155.

Table 1: Size and power (in %) of GARCH and GJR-GARCH tests with $T = 120$

	$\delta = 0.0$				$\delta = 0.5$				$\delta = 1.0$			
	MC	E	LK	DS	MC	E	LK	DS	MC	E	LK	DS
Panel A: GARCH with $N(0,1)$ innovations												
(a_1, b_1)												
$(0.00, 0.80)$	4.7	3.7	4.8	4.8	5.1	3.0	4.2	4.4	4.7	4.1	5.3	5.5
$(0.05, 0.80)$	9.6	9.8	13.0	**13.3**	**13.3**	8.0	11.6	11.7	**23.6**	10.8	15.5	15.7
$(0.05, 0.85)$	12.5	10.1	14.5	**14.8**	**14.1**	9.7	12.4	12.6	**22.4**	10.2	14.5	14.7
$(0.05, 0.89)$	13.5	9.6	13.5	**13.8**	**15.1**	8.6	12.6	12.8	**21.0**	9.1	13.0	13.3
$(0.10, 0.80)$	**28.6**	18.1	24.8	25.4	**36.2**	21.0	27.4	27.6	**55.7**	22.6	29.6	30.4
$(0.10, 0.85)$	**34.7**	22.4	28.4	28.5	**36.6**	19.5	25.6	25.9	**50.0**	20.4	26.1	26.2
$(0.10, 0.89)$	**46.2**	19.7	26.9	27.3	**50.1**	20.7	27.5	27.8	**57.3**	20.1	25.9	26.4
Panel B: GARCH with standardized $t(6)$ innovations												
$(0.00, 0.80)$	4.8	3.1	4.8	5.0	4.6	3.0	4.1	4.2	4.9	3.5	4.7	4.8
$(0.05, 0.80)$	10.0	6.7	10.3	**10.4**	**11.0**	7.3	10.5	10.7	**19.5**	9.3	12.2	12.4
$(0.05, 0.85)$	8.7	7.5	11.0	**11.5**	**10.6**	6.2	9.0	9.2	**16.3**	7.0	10.0	10.1
$(0.05, 0.89)$	**8.8**	6.3	8.7	8.7	**11.4**	6.0	8.1	8.4	**15.6**	7.0	9.3	9.6
$(0.10, 0.80)$	**14.4**	10.7	13.9	14.3	**23.5**	13.7	16.5	16.7	**47.2**	20.0	25.7	26.0
$(0.10, 0.85)$	**18.9**	11.6	15.5	16.1	**24.6**	14.1	18.2	18.2	**44.3**	16.5	21.9	22.3
$(0.10, 0.89)$	**46.1**	22.7	28.0	28.4	**48.0**	20.3	26.5	26.9	**68.9**	24.7	29.1	29.4
Panel C: GJR-GARCH with $N(0,1)$ innovations												
(a_1, γ_1, b_1)												
$(0.00, 0.00, 0.80)$	4.7	3.7	4.3	4.4	5.1	3.5	4.8	5.0	5.0	4.3	5.1	5.1
$(0.05, 0.04, 0.80)$	18.5	14.9	20.9	**21.1**	**22.6**	15.0	21.1	21.6	**31.4**	18.4	22.6	23.0
$(0.05, 0.04, 0.85)$	19.1	13.5	18.9	**19.5**	**22.7**	14.4	18.9	19.2	**31.1**	14.0	18.5	19.2
$(0.05, 0.08, 0.80)$	25.7	18.6	25.7	**25.9**	**33.7**	20.7	27.2	27.7	**48.2**	21.2	27.0	27.6
$(0.05, 0.08, 0.85)$	**28.4**	20.3	25.1	25.4	**35.2**	20.1	26.1	26.9	**52.3**	21.8	29.1	29.4
$(0.10, 0.04, 0.80)$	**37.1**	25.5	33.4	33.9	**43.0**	23.7	30.5	31.2	**60.3**	27.5	34.9	35.0
$(0.10, 0.04, 0.85)$	**49.9**	25.5	34.2	34.8	**53.5**	25.2	33.2	33.7	**66.4**	25.1	32.1	32.6
$(0.10, 0.08, 0.80)$	**45.0**	29.4	35.0	35.2	**59.0**	33.1	40.9	41.9	**70.9**	33.0	41.1	41.4
$(0.10, 0.08, 0.85)$	**69.6**	35.1	42.2	42.3	**73.6**	33.4	41.1	41.9	**80.6**	36.1	43.2	43.5
Panel D: GJR-GARCH with standardized $t(6)$ innovations												
$(0.00, 0.00, 0.80)$	4.8	3.5	4.9	4.9	4.7	3.6	4.1	4.2	5.0	3.4	5.0	5.1
$(0.05, 0.04, 0.80)$	**13.9**	9.5	12.8	13.1	**16.6**	10.8	14.9	15.1	**30.5**	12.6	17.2	17.4
$(0.05, 0.04, 0.85)$	**13.8**	7.7	10.3	10.4	**16.1**	9.8	12.7	12.9	**26.7**	11.7	16.4	16.6
$(0.05, 0.08, 0.80)$	**16.0**	10.7	15.2	15.8	**12.7**	12.0	15.0	15.3	**35.8**	16.8	21.3	21.5
$(0.05, 0.08, 0.85)$	**15.9**	9.6	13.8	13.9	**22.5**	10.7	14.9	15.0	**42.2**	16.9	21.8	22.1
$(0.10, 0.04, 0.80)$	**22.1**	15.3	19.1	19.2	**31.0**	16.1	20.5	20.8	**51.0**	22.7	29.3	29.6
$(0.10, 0.04, 0.85)$	**27.1**	16.1	20.3	20.9	**36.3**	15.0	20.1	20.6	**56.4**	24.1	28.8	29.0
$(0.10, 0.08, 0.80)$	**26.7**	13.5	19.2	19.7	**37.6**	18.0	23.9	24.2	**64.1**	30.5	36.1	36.4
$(0.10, 0.08, 0.85)$	**54.4**	28.4	34.1	34.2	**66.3**	31.5	38.8	39.1	**84.1**	38.5	44.6	45.3

Notes: The DGP is $y_t = \mu + \delta \log(\sigma_t^2) + \varepsilon_t$ with $\varepsilon_t = \sigma_t \eta_t$. The conditional variance process is $\sigma_t^2 = \omega + a_1 \varepsilon_{t-1}^2 + b_1 \sigma_{t-1}^2$ in Panels A and B; and $\sigma_t^2 = \omega + a_1 \varepsilon_{t-1}^2 + \gamma_1 \varepsilon_{t-1}^2 \mathbb{I}[\varepsilon_{t-1} < 0] + b_1 \sigma_{t-1}^2$ in Panels C and D. The null hypothesis of conditional homoskedasticity is represented by $a_1 = 0$ or $a_1 = 0, \gamma_1 = 0$ on the first line in each panel. The results shown under the labels MC, E, LK, and DS correspond to the proposed MC QLR test with $N = 20$, and the LM tests by Engle (1982), Lee and King (1993), and Demos and Sentana (1998), respectively. The nominal level is 5% and the results are based on 1000 replications of each DGP configuration. The entries set in bold show the most powerful tests.

Table 2: Size and power (in %) of GARCH tests

	$\delta = 0.0$				$\delta = 0.5$				$\delta = 1.0$			
(a_1, b_1)	MC	E	LK	DS	MC	E	LK	DS	MC	E	LK	DS
Panel A: $T = 120$, $N(0,1)$ innovations												
$(0.00, 0.80)$	1.7	3.1	4.7	4.8	2.5	4.0	4.6	5.0	2.3	4.3	3.9	5.0
$(0.05, 0.80)$	8.3	8.8	**13.8**	13.5	9.3	11.1	14.0	**14.7**	**14.9**	9.5	12.0	13.9
$(0.05, 0.85)$	9.6	8.7	**12.6**	12.5	9.2	10.2	**14.2**	13.9	**14.6**	10.4	13.1	**14.6**
$(0.05, 0.89)$	8.9	9.1	**13.2**	12.9	11.7	9.7	**13.7**	**13.7**	**17.1**	11.7	15.2	15.1
$(0.10, 0.80)$	23.2	19.8	26.6	**26.7**	**28.3**	21.6	26.2	26.9	**46.2**	23.1	28.4	29.2
$(0.10, 0.85)$	**28.1**	19.3	26.2	26.3	**33.7**	20.6	26.1	27.2	**48.4**	22.1	27.9	27.9
$(0.10, 0.89)$	**43.8**	20.5	26.3	26.5	**43.9**	20.8	27.3	27.3	**44.0**	20.3	26.6	27.7
Panel B: $T = 240$, $N(0,1)$ innovations												
$(0.00, 0.80)$	2.6	4.5	4.6	4.8	2.0	4.3	4.3	4.8	2.1	4.7	5.1	5.2
$(0.05, 0.80)$	9.4	13.7	18.8	**19.1**	15.1	13.7	19.8	**20.1**	23.6	17.7	22.2	**24.1**
$(0.05, 0.85)$	14.1	14.6	19.3	**19.5**	16.6	14.9	20.9	**21.4**	**26.1**	18.4	23.6	24.7
$(0.05, 0.89)$	18.3	15.3	**20.3**	20.2	19.7	16.5	22.1	**22.5**	**27.4**	19.3	25.8	26.0
$(0.10, 0.80)$	40.6	37.1	44.5	**44.8**	**51.7**	37.9	45.6	45.5	**70.5**	41.4	48.6	49.7
$(0.10, 0.85)$	**55.0**	41.1	48.6	48.4	**60.1**	38.4	46.8	47.6	**70.4**	43.1	49.4	49.8
$(0.10, 0.89)$	**74.9**	45.0	52.7	52.8	**76.2**	47.7	55.9	56.2	**75.5**	45.4	52.8	53.5
Panel C: $T = 120$, standardized $t(6)$ innovations												
$(0.00, 0.80)$	2.8	3.5	4.2	4.3	3.5	2.2	3.9	4.2	3.0	4.0	4.9	5.1
$(0.05, 0.80)$	5.8	6.0	8.6	**8.7**	8.6	6.9	10.0	**10.5**	**11.7**	8.2	9.4	11.5
$(0.05, 0.85)$	6.2	6.2	8.4	**8.5**	7.4	6.0	8.8	**8.9**	**12.9**	9.2	11.7	12.7
$(0.05, 0.89)$	7.8	7.5	10.0	**10.2**	8.9	6.8	9.8	**10.1**	**10.8**	7.5	10.7	**10.8**
$(0.10, 0.80)$	12.8	11.7	14.8	**15.2**	**17.9**	11.4	16.1	16.5	**35.8**	20.1	23.7	25.4
$(0.10, 0.85)$	**15.8**	11.6	14.8	15.7	**21.3**	12.5	16.5	17.7	**36.6**	18.8	22.3	24.2
$(0.10, 0.89)$	**42.0**	18.7	22.2	23.3	**47.4**	19.5	23.7	24.6	**65.1**	25.0	28.8	30.5
Panel D: $T = 240$, standardized $t(6)$ innovations												
$(0.00, 0.80)$	2.2	3.0	5.0	5.0	2.4	3.0	3.9	4.0	3.8	3.6	4.6	5.2
$(0.05, 0.80)$	9.8	9.3	12.4	**12.6**	10.5	9.4	12.7	**13.2**	**20.9**	13.7	16.5	17.5
$(0.05, 0.85)$	9.1	8.9	11.3	**11.4**	11.7	10.6	12.7	**12.9**	**20.8**	12.6	16.7	17.1
$(0.05, 0.89)$	10.6	10.3	13.2	**13.3**	12.9	12.4	**15.2**	**15.2**	**21.0**	12.9	16.5	16.8
$(0.10, 0.80)$	20.4	16.9	**23.7**	23.6	**29.3**	21.8	28.2	28.6	**54.8**	33.2	38.3	40.1
$(0.10, 0.85)$	**28.9**	20.6	25.3	25.6	**34.5**	20.1	24.3	24.9	**56.9**	29.7	35.8	36.9
$(0.10, 0.89)$	**60.6**	35.4	40.6	41.7	**65.1**	33.9	39.3	40.4	**79.7**	40.1	45.4	47.1

Notes: The DGP is $y_t = \mu + \delta \log(\sigma_t^2) + \beta x_t + \varepsilon_t$ with $\varepsilon_t = \sigma_t \eta_t$ and $\sigma_t^2 = \omega + a_1 \varepsilon_{t-1}^2 + b_1 \sigma_{t-1}^2$. The null hypothesis of conditional homoskedasticity is represented by $a_1 = 0$ on the first line in each panel. The results shown under the labels MC, E, LK, and DS correspond to the proposed MC QLR test with $N = 20$, and the LM tests by Engle (1982), Lee and King (1993), and Demos and Sentana (1998), respectively. The nominal level is 5% and the results are based on 1000 replications of each DGP configuration. The entries set in bold show the most powerful tests.

Table 3: Size and power (in %) of GJR-GARCH tests

	$\delta = 0.0$				$\delta = 0.5$				$\delta = 1.0$			
(a_1, γ_1, b_1)	MC	E	LK	DS	MC	E	LK	DS	MC	E	LK	DS
Panel A: $T = 120$, $N(0,1)$ innovations												
$(0.00, 0.00, 0.80)$	2.5	4.6	4.2	4.2	1.9	3.7	4.4	4.7	2.7	4.3	4.6	5.3
$(0.05, 0.04, 0.80)$	10.8	12.5	**17.7**	16.9	14.3	12.7	18.2	**18.3**	**23.4**	17.9	22.4	**23.4**
$(0.05, 0.04, 0.85)$	14.8	13.9	**18.5**	18.1	16.8	13.7	17.9	**18.1**	**27.4**	15.9	19.8	20.4
$(0.05, 0.08, 0.80)$	19.5	20.5	**26.2**	26.0	**25.4**	19.0	25.1	**25.4**	**39.2**	19.6	25.4	26.7
$(0.05, 0.08, 0.85)$	21.6	18.1	**23.8**	23.4	**27.3**	18.6	23.9	23.9	**40.7**	20.7	25.9	26.4
$(0.10, 0.04, 0.80)$	27.0	24.0	**30.6**	30.3	**36.1**	28.0	34.1	34.6	**56.0**	27.5	33.2	34.1
$(0.10, 0.04, 0.85)$	**40.3**	25.9	32.0	32.1	**48.6**	25.1	29.9	30.5	**63.1**	27.0	33.5	33.7
$(0.10, 0.08, 0.80)$	**41.7**	31.7	39.6	39.6	**47.2**	29.9	37.4	37.7	**66.9**	33.0	38.5	39.8
$(0.10, 0.08, 0.85)$	**61.7**	31.6	39.4	39.6	**70.6**	33.4	40.4	41.0	**77.4**	32.1	39.1	39.8
Panel B: $T = 240$, $N(0,1)$ innovations												
$(0.00, 0.00, 0.80)$	2.6	3.7	5.1	5.1	2.3	2.8	4.9	5.1	2.0	4.5	4.9	5.0
$(0.05, 0.04, 0.80)$	16.7	23.3	**29.9**	29.7	24.6	25.5	31.5	**31.8**	**38.3**	28.0	34.4	35.9
$(0.05, 0.04, 0.85)$	22.7	26.8	33.3	**33.4**	26.1	24.3	31.5	**31.9**	**42.0**	25.9	32.7	33.5
$(0.05, 0.08, 0.80)$	29.9	30.8	39.5	**39.7**	40.2	34.8	42.5	**42.9**	**62.0**	39.1	46.6	47.1
$(0.05, 0.08, 0.85)$	42.5	35.9	42.5	**42.6**	**49.5**	36.1	43.5	43.8	**65.2**	37.2	46.3	46.9
$(0.10, 0.04, 0.80)$	46.2	46.1	54.1	**54.4**	**58.9**	48.3	56.2	56.7	**79.2**	51.4	59.4	60.0
$(0.10, 0.04, 0.85)$	**71.8**	52.1	59.7	60.3	**74.2**	51.1	57.8	58.4	**86.4**	51.3	57.7	58.2
$(0.10, 0.08, 0.80)$	**67.2**	57.1	64.8	65.2	**74.3**	55.4	63.0	63.3	**88.4**	59.9	66.2	66.7
$(0.10, 0.08, 0.85)$	**94.3**	67.8	74.2	74.8	**94.0**	65.7	73.3	73.4	**97.2**	64.4	70.8	71.5
Panel C: $T = 120$, standardized $t(6)$ innovations												
$(0.00, 0.00, 0.80)$	2.4	3.8	5.0	5.0	3.1	3.4	4.8	5.1	2.7	3.7	4.0	5.1
$(0.05, 0.04, 0.80)$	9.3	9.0	12.0	**12.1**	11.7	9.7	13.3	**13.6**	**19.2**	11.8	14.5	15.9
$(0.05, 0.04, 0.85)$	7.7	7.4	9.6	**9.9**	**12.2**	8.1	11.5	11.7	**19.6**	12.5	15.1	16.7
$(0.05, 0.08, 0.80)$	11.2	9.2	12.5	**12.7**	15.0	11.0	14.8	**15.8**	**30.0**	17.1	20.5	22.1
$(0.05, 0.08, 0.85)$	10.8	10.8	**14.0**	**14.0**	**16.6**	10.9	14.9	15.2	**27.9**	13.7	17.8	18.8
$(0.10, 0.04, 0.80)$	17.8	15.2	18.8	**19.5**	21.1	16.5	20.8	**21.6**	**42.8**	21.3	26.2	27.5
$(0.10, 0.04, 0.85)$	**24.0**	16.3	20.3	20.9	**28.2**	16.3	21.3	21.8	**47.1**	20.9	25.0	26.2
$(0.10, 0.08, 0.80)$	**20.7**	16.2	19.8	20.2	**29.5**	16.9	21.8	23.1	**52.8**	25.5	30.3	32.4
$(0.10, 0.08, 0.85)$	**52.4**	27.4	32.5	34.4	**57.7**	31.0	36.0	37.8	**79.8**	39.7	42.7	45.0
Panel D: $T = 240$, standardized $t(6)$ innovations												
$(0.00, 0.00, 0.80)$	3.6	3.7	4.8	4.8	3.5	4.0	4.9	5.0	4.1	3.8	4.7	4.9
$(0.05, 0.04, 0.80)$	12.3	11.4	14.9	**15.0**	17.8	14.3	18.8	**19.3**	**32.4**	24.1	27.6	28.7
$(0.05, 0.04, 0.85)$	13.3	11.2	**15.5**	**15.5**	16.4	11.9	16.3	**17.0**	**31.5**	16.0	20.7	21.2
$(0.05, 0.08, 0.80)$	16.5	15.3	19.1	**19.2**	26.5	20.6	26.4	**26.7**	**47.4**	29.2	33.2	35.1
$(0.05, 0.08, 0.85)$	19.9	17.7	24.1	**24.1**	**26.5**	19.5	24.5	24.8	**47.7**	28.5	32.6	34.0
$(0.10, 0.04, 0.80)$	22.3	19.7	**24.8**	**24.8**	**35.1**	25.0	31.3	31.8	**66.6**	40.8	46.9	48.0
$(0.10, 0.04, 0.85)$	**33.8**	29.2	33.7	33.6	**44.6**	29.3	35.4	36.4	**68.5**	36.4	42.4	43.9
$(0.10, 0.08, 0.80)$	**35.0**	27.4	33.1	33.4	**47.9**	31.1	37.8	38.8	**74.5**	44.8	51.2	53.0
$(0.10, 0.08, 0.85)$	**63.3**	44.6	50.9	52.0	**74.4**	49.1	54.2	55.8	**91.8**	58.2	63.2	64.7

Notes: The DGP is $y_t = \mu + \delta \log(\sigma_t^2) + \beta x_t + \varepsilon_t$ with $\varepsilon_t = \sigma_t \eta_t$ and $\sigma_t^2 = \omega + a_1 \varepsilon_{t-1}^2 + \gamma_1 \varepsilon_{t-1}^2 \mathbb{I}[\varepsilon_{t-1} < 0] + b_1 \sigma_{t-1}^2$. The null hypothesis is represented by $a_1 = 0, \gamma_1 = 0$ on the first line in each panel. The results shown under the labels MC, E, LK, and DS correspond to the proposed MC QLR test with $N = 20$, and the LM tests by Engle (1982), Lee and King (1993), and Demos and Sentana (1998), respectively. The nominal level is 5% and the results are based on 1000 replications of each DGP configuration. The entries set in bold show the most powerful tests.

Table 4: Tests for GARCH effects in Fama-French returns

	Lo 10	Dec 2	Dec 3	Dec 4	Dec 5	Dec 6	Dec 7	Dec 8	Dec 9	Hi 10
Panel A: Basic specification, Jan 2001–Dec 2010 ($T = 120$)										
MC(GARCH, $\delta = 0$)	0.01	**0.02**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MC(GARCH, δ free)	0.01	**0.01**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MC(GJR-GARCH, $\delta = 0$)	0.01	**0.01**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MC(GJR-GARCH, δ free)	0.01	**0.01**	0.02	0.01	0.01	0.01	0.01	0.01	0.01	0.01
E	0.00	**0.34**	0.09	0.08	0.00	0.00	0.00	0.00	0.00	0.00
LK	0.00	**0.17**	0.05	0.04	0.00	0.00	0.00	0.00	0.00	0.00
DS	0.00	**0.17**	0.05	0.04	0.00	0.00	0.00	0.00	0.00	0.00
Panel B: Basic specification, Jan 1991–Dec 2010 ($T = 240$)										
MC(GARCH, $\delta = 0$)	0.01	**0.01**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MC(GARCH, δ free)	0.01	**0.01**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MC(GJR-GARCH, $\delta = 0$)	0.01	**0.01**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
MC(GJR-GARCH, δ free)	0.01	**0.01**	0.01	0.01	0.01	0.01	0.01	0.01	0.01	0.01
E	0.00	**0.16**	0.03	0.01	0.00	0.00	0.00	0.00	0.00	0.00
LK	0.00	**0.08**	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
DS	0.00	**0.08**	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Panel C: CAPM-type specification, Jan 2001–Dec 2010 ($T = 120$)										
MC(GARCH, $\delta = 0$)	0.02	0.01	0.01	0.01	0.03	0.27	0.46	0.16	0.05	0.19
MC(GARCH, δ free)	0.01	0.01	0.01	0.01	0.01	0.26	0.77	0.14	0.02	0.18
MC(GJR-GARCH, $\delta = 0$)	0.04	0.01	0.01	0.01	0.03	0.27	0.40	0.02	0.01	0.26
MC(GJR-GARCH, δ free)	0.01	0.01	0.01	0.03	0.02	0.31	0.88	0.35	0.02	0.19
E	0.00	0.05	0.02	0.08	0.03	0.85	0.83	0.72	0.10	0.56
LK	0.00	0.05	0.01	0.04	0.01	0.41	0.59	0.35	0.05	0.27
DS	0.00	0.03	0.00	0.04	0.01	0.42	0.50	0.36	0.05	0.28
Panel D: CAPM-type specification, Jan 1991–Dec 2010 ($T = 240$)										
MC(GARCH, $\delta = 0$)	0.01	0.01	0.01	0.01	0.01	**0.01**	**0.17**	0.21	**0.01**	0.01
MC(GARCH, δ free)	0.01	0.01	0.01	0.01	0.04	**0.01**	**0.01**	0.31	**0.01**	0.01
MC(GJR-GARCH, $\delta = 0$)	0.01	0.01	0.01	0.01	0.01	**0.01**	**0.01**	0.34	**0.01**	0.01
MC(GJR-GARCH, δ free)	0.01	0.01	0.01	0.01	0.01	**0.13**	**0.01**	0.21	**0.01**	0.01
E	0.00	0.00	0.00	0.00	0.00	**0.19**	**0.65**	0.79	**0.41**	0.00
LK	0.00	0.00	0.00	0.00	0.00	**0.09**	**0.32**	0.39	**0.20**	0.00
DS	0.00	0.00	0.00	0.00	0.00	**0.09**	**0.32**	0.40	**0.20**	0.00

Notes: The entries are p-values for the proposed MC test with $N = 100$ assuming a symmetric GARCH(1,1) and an asymmetric GJR-GARCH(1,1), and for the Engle (1982) (E), Lee and King (1993) (LK), and Demos and Sentana (1998) (DS) tests. The data are monthly excess returns of 10 portfolios formed on size for the period from January 2001 to December 2010 (Panels A and C) and from January 1991 to December 2010 (Panels B and D). The results are organized from the smallest decile (Lo 10) to the highest one (Hi 10). For each decile, the basic specification is given by (20) and the CAPM-type specification corresponds to (21) with the market risk factor as x_t. Those two specifications are also examined when δ is set to zero and when it is a free parameter. The columns set in bold show important disagreements between the E-LK-DS group and the MC tests. With $N = 100$, the smallest possible MC p-value is 1%.