A diagnostic m-test for distributional specification of parametric conditional heteroscedasticity models for financial data

Bernard Lejeune[1]

University of Liège, CORE and ERUDITE

Boulevard du Rectorat, 7, B33
4000 Liège Belgium
E-mail: B.Lejeune@ulg.ac.be

April 2002

Abstract

This paper proposes a convenient and generally applicable diagnostic m-test for checking the distributional specification of parametric conditional heteroscedasticity models for financial data such as the customary Student t GARCH model. The proposed test is based on the moments of the probability integral transform of the innovations of the assumed model. Monte-Carlo evidence indicates that our suggested test performs well both in terms of size and power.

Keywords: Parametric conditional heteroscedasticity models, distributional specification test, m-testing.
J.E.L. classification: C12, C22, C52.

[1] I wish to thank Luc Bauwens for helpful comments on an earlier version of this paper.

1. Introduction

Parametric conditional heteroscedasticity models, as for example the standard Student t GARCH model of Bollerslev (1987), are a well-established and indisputably fruitful tool for analyzing …nancial data. If these models are routinely estimated in empirical applications, diagnostic testing of their speci…cation, and more particularly of their distributional speci…cation, is however much less common in practice. Intended for helping to …ll this gap, this paper proposes a convenient and generally applicable moment-based diagnostic test allowing to readily check the distributional aspect of these models.

The …rst idea which naturally comes in mind when thinking about a diagnostic test for checking the distributional speci…cation of parametric conditional heteroscedasticity models is, by analogy to the popular Jarque-Bera (1980) test for normality in standard homoscedastic regression models, to check through a m-test (Newey (1985), Tauchen (1985), White (1987, 1994)) that the third and fourth order sample moments of the (estimated) innovations of the model are not signi…cantly di¤erent from their (estimated) theoretical values as they follow from the assumed distribution for the innovations.

This way of looking at distributional misspeci…cation is both intuitively appealing and convenient since m-tests are standard and easy to implement. Unfortunately, it is not always applicable, in particular when dealing with a number of popular models as for example the already mentioned standard Student t GARCH model.

Indeed, as a general rule, to be applicable, any m-test of the null hypothesis that the r-th moment of some quantity is equal to a given value requires that, under the null, at least the 2r-th moment of this quantity be …nite. Accordingly, the above testing strategy can only be applied provided that the distribution of the innovations underlying of the model speci…cation possesses, under the null of correct speci…cation, …nite moments at least up to order eight. If there is obviously nothing to worry about regarding this moment condition whenever working with distributions which always possess …nite moments of all orders, this is no longer true when dealing with distributions which in practice may turn out to possess only few …nite moments.

As a matter of fact, empirical studies using the standard Student t(º) GARCH model, whose by de…nition underlying Student t(º) innovations only possess …nite moments up to order ± < º, typically report estimated values between 4 and 8 for the number of degrees of freedom º, suggesting that the underlying innovations usually do not possess enough …nite moments for the above Jarque-Bera like moment-based test of the Student t distributional assumption to be applicable. The same problem is likely to arise whenever working with parametric conditional heteroscedasticity models relying on fat-tail distributions which, as the Student t, do not necessarily possess …nite moments of all orders. This includes, among others, the generalized Student t GARCH model of Bollerslev and al. (1994), the skewed Student t GARCH models of Hansen (1994) and Lambert and al. (2001), or the generalized Student t GARCH model as considered in Lye and al. (1998).

To circumvent this problem while staying in the convenient m-testing framework, this paper proposes an alternative diagnostic m-test consisting in checking, instead of the sample moments of the (estimated) innovations, the sample moments of the probability integral transform (i.e. the cdf. transform corresponding to the assumed pdf. for the innovations) of the (estimated) innovations, which, if the assumed pdf. for the innovations is correct and regardless of the precise form of this postulated pdf., should not be significantly different from the moments (by definition all finite) of the uniform distribution on $[0;1]$.

Unlike the "direct" Jarque-Bera like moment-based diagnostic test outline above, this "indirect" moment-based way of looking at distributional misspecification, which is in the line of the graphical and informal procedure proposed by Diebold and al. (1998) to evaluate density forecasts, is by construction applicable regardless whether or not the distribution of the innovations underlying the model possesses, under the null of correct specification, only few first finite moments, providing thereby a convenient diagnostic test of general applicability.

The rest of the paper is organized as follows. Section 2 briefly describes the class of parametric conditional heteroscedasticity models we consider and their maximum likelihood estimation. Section 3 develops the rationale of our proposed m-testing strategy for checking the distributional specification of these models. Section 4 suggests two alternative proper m-test statistics for implementing it. Section 5 presents some Monte-Carlo evidence on the performance of the different possible versions of the proposed test. Finally, Section 6 concludes.

2. Model specification and estimation

Let y_t stand for the dependent variable of interest, which is assumed to be continuous, z_t designate some $\tau \times 1$ vector of explanatory variables allowed to be continuous, discrete or mixed, and x_t denote the information set $x_t \equiv (z_t; \psi_{t-1})$, where $\psi_{t-1} \equiv (y_{t-1}; z_{t-1}; \ldots; y_1; z_1)$ is the information available on y and z at time $t-1$. If there is no explanatory variables z_t; x_t is reduced to the information available on y at time $t-1$, i.e. to the information set $x_t \equiv (y_{t-1}; \ldots; y_1)$.

We suppose that (some special case of) the following generic parametric conditional heteroscedasticity model is considered for modelling y_t in terms of x_t:

$$y_t = \mu_t(x_t; \gamma) + \sqrt{h_t(x_t; \gamma)}\, \varepsilon_t; \quad t = 1; 2; \ldots$$

where γ is a $k_\gamma \times 1$ vector of parameters, the functions $\mu_t(.;.)$ and $h_t(.;.) > 0$ are known scalar functions, and ε_t are zero mean and unit variance innovations independent of x_t and identically and independently distributed with density $g(\varepsilon; \eta)$, where η is a $k_\eta \times 1$ vector of shape parameters.

The above specification defines a fully parametric model P for the conditional densities of y_t given x_t:

$$P \equiv \left\{ f_t(y_t|x_t; \theta) = \frac{1}{\sqrt{h_t(x_t; \gamma)}}\, g\left(\frac{y_t - \mu_t(x_t; \gamma)}{\sqrt{h_t(x_t; \gamma)}}; \eta \right) : \theta = (\gamma'; \eta')' \in \Theta \right\}; \quad t = 1; 2; \ldots$$

whose, by construction, first two conditional moments of y_t given x_t are

$$E(y_t|x_t) = \mu_t(x_t;\gamma) \text{ and } V(y_t|x_t) = h_t(x_t;\gamma);\quad t = 1, 2, ...$$

This parametric model generically describes the most commonly used class of parametric models for modelling financial data. In practical applications, $\mu_t(x_t;\gamma)$ is typically specified according to an AR, MA or ARMA process, possibly including contemporaneous and lagged values of some explanatory variables z as well as some function of the conditional variance (in ARCH-M type models), and $h_t(x_t;\gamma)$ according to some autoregressive scheme such as ARCH, GARCH, EGARCH, etc... On its side, the density $g(\varepsilon;\eta)$ of the innovations ε_t is chosen among standardized (i.e. transformed such that they have zero mean and unit variance, regardless of the value of their shape parameter(s) η) continuous distributions allowing for fatter tails than the normal distribution and possibly further for asymmetry[2]. A customary example of such models is the pure time-series (i.e. without explanatory variables z) Student $t(\nu)$ AR(1) - GARCH(1,1) model obtained by setting

$$\mu_t(x_t;\gamma) = \gamma_1 + \gamma_2 y_{t-1};\quad h_t(x_t;\gamma) = \gamma_3 + \gamma_4 u_{t-1}^2 + \gamma_5 h_{t-1} \tag{1}$$

and

$$g(\varepsilon;\eta) = b g^*(b\varepsilon;\nu) \tag{2}$$

where $\gamma = (\gamma_1, ..., \gamma_5)'$, $\eta = \nu > 2$, $u_{t-1} = y_{t-1} - \mu_{t-1}(x_{t-1};\gamma)$, $h_{t-1} = h_{t-1}(x_{t-1};\gamma)$, $b = \sqrt{\frac{\nu}{\nu-2}}$ and $g^*(w;\nu)$ denotes the usual (i.e. non-standardized) Student t density with ν degrees of freedom[3].

A natural estimator of the parameters of model P is given by the ML estimator

$$\hat{\theta}_n = (\hat{\gamma}_n', \hat{\eta}_n')' = \text{Argmax}_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} l_t(y_t;x_t;\theta) \tag{3}$$

where

$$l_t(y_t;x_t;\theta) = -0.5 \ln h_t(x_t;\gamma) + \ln g\left(\frac{y_t - \mu_t(x_t;\gamma)}{\sqrt{h_t(x_t;\gamma)}};\eta\right)$$

Under general regularity conditions (e.g. White (1994) or Wooldridge (1994)), if model P is correctly specified, i.e. if there exists some true value θ^o in Θ such that

$$f_t(y_t|x_t;\theta^o) = p_t^o(y_t|x_t);\quad t = 1, 2, ...$$

where $p_t^o(y_t|x_t)$ denotes the true conditional density of y_t given x_t, the ML estimator (3) yields a consistent and efficient estimator of the unknown $k \times 1$ true value $\theta^o = (\gamma^{o\prime}, \eta^{o\prime})'$ of P, whose limiting distribution is

$$V_n^{o\,-\frac{1}{2}}\sqrt{n}\left(\hat{\theta}_n - \theta^o\right) \xrightarrow{d} N(0, I_k)$$

[2] For a general survey on these models, see Bollerslev and al. (1994) or Palm (1996). For a more specific account about available specifications for the density of the innovations, see Paolella (1999).

[3] i.e. $g^*(w;\nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\,\Gamma\left(\frac{\nu}{2}\right)}\left(1 + \frac{w^2}{\nu}\right)^{-\frac{\nu+1}{2}}$.

where

$$V_n^o = -\frac{1}{n}\sum_{t=1}^{n} E\left[H_t^o\right] = \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^o s_t^{o\prime}\right]$$

with

$$s_t^o = \frac{\partial l_t(y_t; x_t; \theta^o)}{\partial \theta} \quad \text{and} \quad H_t^o = \frac{\partial^2 l_t(y_t; x_t; \theta^o)}{\partial \theta \partial \theta'} \tag{4}$$

3. Testing distributional specification through moments of probability integral transform

In what follows, we suppose that interest lies in checking the distributional specification of the tentatively postulated parametric model P, i.e. in testing the null hypothesis that model P is correctly specified against the alternative that it is not due to misspecification of the assumed density $g(\varepsilon; \eta)$ for the innovations, the correctness of the specification of the conditional mean and conditional variance being taken for granted. In practice, this latter point may for example be relevantly and comprehensively checked using the convenient Wooldridge's robust LM-type diagnostic m-tests based on auxiliary (Gaussian) pseudo-maximum likelihood estimation of the model (see Wooldridge (1990, 1991a, 1991b) and Bollerslev and al. (1992)).

As outlined in the introduction, by analogy to the popular Jarque-Bera (1980) test for normality in standard homoscedastic regression models, at first sight, a natural strategy for testing the distributional specification of model P would be to check through a m-test that the misspecification indicator

$$\hat{M}_n = \frac{1}{n}\sum_{t=1}^{n} \begin{pmatrix} \hat{e}_t^3 - \phi_3(\hat{\eta}_n) \\ \hat{e}_t^4 - \phi_4(\hat{\eta}_n) \end{pmatrix} \tag{5}$$

where

$$\hat{e}_t = e_t(y_t; x_t; \hat{\gamma}_n) = \frac{y_t - \mu_t(x_t; \hat{\gamma}_n)}{\sqrt{h_t(x_t; \hat{\gamma}_n)}} \quad \text{and} \quad \phi_r(\eta) = \int_{-\infty}^{+\infty} \varepsilon^r g(\varepsilon; \eta) d\varepsilon$$

is not significantly different from zero.

The rationale of such a testing strategy stems from the fact that under correct specification of P, the ML estimator $\hat{\theta}_n = (\hat{\gamma}_n', \hat{\eta}_n')'$ is consistent for $\theta^o = (\gamma^{o\prime}, \eta^{o\prime})'$ and for any r such that $\phi_r(\eta^o)$ is finite

$$E\left[(e_t^o)^r - \phi_r(\eta^o)\right] = 0; \quad t = 1, 2, \ldots \tag{6}$$

where $e_t^o = e_t(y_t; x_t; \gamma^o)$, while under distributional misspecification of P, $\hat{\theta}_n$ is consistent for some pseudo-true value $\theta_n^* = (\gamma_n^{*\prime}, \eta_n^{*\prime})'$ (generally) different from θ^o and (again generally)

$$E\left[(e_t^*)^r - \phi_r(\eta_n^*)\right] \neq 0; \quad t = 1, 2, \ldots$$

where $e_t^* = e_t(y_t; x_t; \gamma_n^*)$. Further, it follows from the observation that among the possible moment restrictions to check suggested by (6), the ones corresponding to the third and the fourth order moments (i.e. skewness and kurtosis) of the innovations appear to be the most relevant to consider[4].

For such a testing strategy to be applicable, it is however necessary that, under the null of correct specification, at the true parameter value $\theta^o = (\gamma^{o\prime}; \eta^{o\prime})^\prime$, the density $g(\varepsilon;\eta^o)$ possesses at least finite first eight order moments (this is required for being able to invoke the central limit theorem on which relies a m-test of the closeness to zero of (5)). As we already argued, it is beyond what we can expect to be fulfilled in typical empirical applications when working with a number popular models[5].

As a general device to circumvent this problem while staying in the convenient m-testing framework, following the line of the informal procedure proposed by Diebold and al. (1998) for evaluating density forecasts, we thus propose to consider the following alternative strategy consisting in checking, instead of the sample moments of the estimated innovations, the sample moments of the probability integral transform (i.e. the cdf. transform corresponding to the assumed pdf. $g(\varepsilon;\eta)$) of the estimated innovations $\hat e_t$ of the model.

The probability integral transform of the estimated innovations $\hat e_t$ of the model are given by

$$\hat v_t = v_t(y_t; x_t; \hat\theta_n) = G(e_t(y_t; x_t; \hat\gamma_n); \hat\eta_n) = G(\hat e_t; \hat\eta_n)$$

where

$$G(\varepsilon;\eta) = \int_{-\infty}^{\varepsilon} g(w;\eta)dw$$

is the cdf. associated to the density $g(\varepsilon;\eta)$.

Following Diebold and al. (1998), it is easily seen that, whatever the precise form of the assumed density $g(\varepsilon;\eta)$, if model P is correctly specified, $v_t^o = v_t(y_t; x_t; \theta^o) = G(e_t^o; \eta^o)$ must be independent of x_t and identically and independently distributed according to a continuous uniform r.v. over $[0;1]$, which by definition possesses finite moments of all orders.

The central moments of a continuous uniform r.v. over $[0;1]$ being, according to Johnson and al. (1970), given by

$$\delta(r) = \begin{cases} \frac{1}{2^r(r+1)} & \text{if } r \text{ is even} \\ 0 & \text{if } r \text{ is odd} \end{cases}$$

[4] This does not mean that moment restrictions corresponding to other moments are not worth to be considered. Our point is simply that, if we want the test to have reasonable power, the set of chosen moment restrictions on which it is based should at least embody moments restrictions corresponding to these two moments.

[5] Note in contrast that for the suggested above Wooldridge's robust diagnostic m-tests of the conditional mean and conditional variance specifications to be applicable, it is enough that, under the null of correct specification, the innovations possess finite first four — first two if only conditional mean diagnostic m-tests are considered — order moments, a condition which is much less stringent and, as suggested by usual empirical results, likely to be fulfilled in most practical applications.

for any $r = 1, 2, \ldots$, under correct specification of P, we must have

$$E\left[(v_t^o - 0.5)^r - \delta(r)\right] = 0, \quad t = 1, 2, \ldots$$

while under distributional misspecification, similarly to above, we will (generally) have

$$E\left[(v_t^* - 0.5)^r - \delta(r)\right] \neq 0, \quad t = 1, 2, \ldots$$

where $v_t^* = v_t(y_t; x_t; \theta_n^*) = G(e_t^*; \eta_n^*)$.

This suggests considering to check the distributional specification of model P by testing through a m-test the closeness to zero of a misspecification indicator of the form

$$\hat{M}_n = \frac{1}{n}\sum_{t=1}^{n} m_t(y_t; x_t; \hat{\theta}_n); \text{ where } m_t(y_t; x_t; \hat{\theta}_n) = \begin{bmatrix} \hat{v}_t - 0.5 \\ (\hat{v}_t - 0.5)^2 - \delta(2) \\ \vdots \\ (\hat{v}_t - 0.5)^q - \delta(q) \end{bmatrix} \qquad (7)$$

for some integer q.

Contrary to the "direct" Jarque-Bera like moment-based strategy, this "indirect" strategy is applicable regardless whether or not, under the null of correct specification, at the true parameter value $\theta^o = (\gamma^{o\prime}; \eta^{o\prime})'$, the density $g(\varepsilon; \eta^o)$ possesses only few first finite moments. And because by definition v_t^o possesses finite moments of all orders, it is also applicable without restriction for any choice of q. It is thus of general applicability.

Theoretically, setting $q = 2$, i.e. checking only the first two order sample moments of $\hat{v}_t$, already allows to detect departures from the assumed density for the innovations both in terms of skewness and kurtosis. Nevertheless, for ensuring power against a large spectrum of alternatives, it is likely that setting $q = 4$ (or possibly further $q = 6$) is a better practical choice.

4. Test statistics

Using the general results of White (1987, 1994), it may be verified that, given the characteristics of the assumed statistical setup, a proper m-test statistic for checking the closeness to zero of the $q \times 1$ misspecification indicator (7) is, under general regularity conditions, given by the asymptotically chi-square statistic

$$M_n = n\hat{M}_n' \hat{K}_n^{-1} \hat{M}_n \overset{d}{\to} \chi^2(q) \qquad (8)$$

where $\hat{K}_n$ stands for any consistent estimator of

$$K_n^o = \frac{1}{n}\sum_{t=1}^{n} E\left[\left(m_t^o - D_n^o A_n^{o-1} s_t^o\right)\left(m_t^o - D_n^o A_n^{o-1} s_t^o\right)'\right] \qquad (9)$$

$$= \frac{1}{n}\sum_{t=1}^{n} E[m_t^o m_t^{o\prime}] - \frac{1}{n}\sum_{t=1}^{n} E[m_t^o s_t^{o\prime}]\left(\frac{1}{n}\sum_{t=1}^{n} E[s_t^o s_t^{o\prime}]\right)^{-1} \frac{1}{n}\sum_{t=1}^{n} E[s_t^o m_t^{o\prime}] \qquad (10)$$

where

$$m_t^o = m_t(y_t; x_t; \theta^o); \quad A_n^o = \frac{1}{n}\sum_{t=1}^{n} E\,[H_t^o]; \quad D_n^o = \frac{1}{n}\sum_{t=1}^{n} E\left[\frac{\partial m_t(y_t; x_t; \theta^o)}{\partial \theta'}\right]$$

s_t^o and H_t^o are as defined in (4), and the equality of (9) and (10) follows from the so-called information matrix (i.e. $A_n^o = -\frac{1}{n}\sum_{t=1}^{n} E\,[s_t^o s_t^{o\prime}]$) and cross-information matrix ($D_n^o = -\frac{1}{n}\sum_{t=1}^{n} E\,[m_t^o s_t^{o\prime}]$) equalities.

The simplest operational form of (8) is obtained by taking as a consistent estimator of K_n^o the empirical counterpart of (10)

$$\hat{K}_n^{OPG} = \frac{1}{n}\sum_{t=1}^{n} \hat{m}_t \hat{m}_t' - \frac{1}{n}\sum_{t=1}^{n} \hat{m}_t \hat{s}_t' \left(\frac{1}{n}\sum_{t=1}^{n} \hat{s}_t \hat{s}_t'\right)^{-1} \frac{1}{n}\sum_{t=1}^{n} \hat{s}_t \hat{m}_t'$$

where

$$\hat{m}_t = m_t(y_t; x_t; \hat{\theta}_n) \quad \text{and} \quad \hat{s}_t = \frac{\partial l_t(y_t; x_t; \hat{\theta}_n)}{\partial \theta}$$

This yields the standard so-called outer-product-gradient (hence the "OPG" superscript) m-test statistic

$$M_n^{OPG} = n\hat{M}_n \left(\hat{K}_n^{OPG}\right)^{-1} \hat{M}_n$$

which in practice may be computed as n minus the residual sum of squares ($= nR_u^2$, R_u^2 denoting the uncentered R-squared) of the OLS artificial regression

$$1 = [\hat{m}_t' \vdots \hat{s}_t']\, b + \text{residuals}, \quad t = 1; 2; \ldots; n$$

where b is a $(q + k_\gamma + k_\eta) \times 1$ vector of auxiliary parameters.

This standard m-test statistic is very easy to implement. It is however well-known[6] for often exhibiting poor finite sample properties: m-tests — and particularly m-tests of high order moments — based on this statistic typically tend to over-reject the null when it is true, i.e. to have an actual finite sample size higher than their nominal asymptotic size, and that even in quite large samples.

According to our experience, as a general rule, a usually better behaved in finite sample version of (8) is given by the asymptotically equivalent statistic (entitled "PML" because it is the standard form of m-test statistics when working in a pseudo-maximum likelihood framework)

$$M_n^{PML} = n\hat{M}_n \left(\hat{K}_n^{PML}\right)^{-1} \hat{M}_n$$

constructed using as a consistent estimator[7] of K_n^o, instead of the empirical coun-

[6] See Davidson and al. (1993), Ch 16.

[7] Note that both $\hat{K}_n^{OPG}$ and $\hat{K}_n^{PML}$ are always at least semi-positive definite (and usually positive definite), ensuring thereby that the calculated statistic never turns out to be negative. This is not the case of other conceivable straightforward consistent estimators of K_n^o such as $\hat{K}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{m}_t\hat{m}_t' + \hat{D}_n \hat{A}_n^{-1} \hat{D}_n'$, $\hat{K}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{m}_t\hat{m}_t' + \left(\frac{1}{n}\sum_{t=1}^{n} \hat{m}_t\hat{s}_t'\right) \hat{A}_n^{-1} \left(\frac{1}{n}\sum_{t=1}^{n} \hat{s}_t\hat{m}_t'\right)$ or $\hat{K}_n = \frac{1}{n}\sum_{t=1}^{n} \hat{m}_t\hat{m}_t' - \hat{D}_n \left(\frac{1}{n}\sum_{t=1}^{n} \hat{s}_t\hat{s}_t'\right)^{-1} \hat{D}_n'$,

terpart of (10), the empirical counterpart of (9)

$$\hat{K}_n^{PML} = \frac{1}{n}\sum_{t=1}\left(\hat{m}_t - \hat{D}_n\hat{A}_n^{-1}\hat{s}_t\right)\left(\hat{m}_t - \hat{D}_n\hat{A}_n^{-1}\hat{s}_t\right)'$$

where $\hat{m}_t$ and $\hat{s}_t$ are as defined above,

$$\hat{D}_n = \frac{1}{n}\sum_{t=1}\frac{\partial m_t(y_t;x_t;\hat{\theta}_n)}{\partial\theta'} \quad \text{and} \quad \hat{A}_n = \frac{1}{n}\sum_{t=1}\frac{\partial^2 l_t(y_t;x_t;\hat{\theta}_n)}{\partial\theta\partial\theta'}$$

Interestingly, this alternative statistic may in practice also be computed through an OLS artificial regression, namely as n minus the residual sum of squares ($= nR_u^2$) of the OLS regression

$$1 = \left[\hat{m}_t' - \hat{s}_t'\hat{A}_n^{-1}\hat{D}_n'\right] b + \text{residuals}, \quad t = 1, 2, \ldots, n$$

where b is here a $q \times 1$ vector of auxiliary parameters. However, because it requires the additional calculation[8] of the matrix of derivatives $\hat{D}_n$ and $\hat{A}_n$, while still convenient, it is not as computationally easy as M_n^{OPG}.

5. Monte-Carlo evidence

Is our proposed distributional diagnostic m-testing strategy effective? Which m-test statistic and how much moment restrictions should we actually use in practice for implementing it? To get insights about these questions, we performed some simulation experiments.

In this simulation study, we investigated the finite sample performance of six versions of our proposed test, namely its M_n^{OPG} and M_n^{PML} forms with $q = 2$, $q = 4$ and $q = 6$, for checking the distributional specification of two models: on one hand, the customary Student $t(\nu)$ AR(1)-GARCH(1,1) model (1)-(2) outlined in Section 2 (hereafter denoted "Model 1"), and on the other hand, a distributional extension of this AR(1)-GARCH(1,1) model where the innovations ε_t are assumed to be distributed according to the standardized form of the skewed Student $t(\nu;\kappa)$ distribution of Fernandez and al. (1998)[9], whose density $g(\varepsilon;\eta)$ may be written

$$g(\varepsilon;\eta) = \frac{2b\kappa}{\kappa^2+1}\left[I_{(a+b\varepsilon<0)}\,g^*(\kappa(a+b\varepsilon);\nu) + I_{(a+b\varepsilon\geq 0)}\,g^*\left(\frac{a+b\varepsilon}{\kappa};\nu\right)\right] \tag{11}$$

---

a characteristic which a priori discards them as relevant candidates for constructing an operational version of (8).

[8] In practice, this may simply be done numerically. In this respect, a Gauss procedure automatically computing, for any choice of q and on the basis of the maximum likelihood estimate $\hat{\theta}_n$, the log-likelihood function $l_t(y_t;x_t;\theta)$ and the probability integral transform function $v_t(y_t;x_t;\theta)$ of the model, both the M_n^{OPG} and M_n^{PML} forms of our proposed distributional diagnostic m-test may be obtained upon request from the author. To run properly, it requires Gauss for Windows 3.2 or above.

[9] The use of this skewed Student $t(\nu;\kappa)$ distribution in the context of parametric GARCH models has recently been considered and advocated in Lambert and al. (2001) and Giot and al. (2001).

where $\theta = (\nu;\xi)'$ with $\nu > 2$ and $\xi > 0$, while $I_{(\cdot)}$ is a 0-1 indicator function,

$$a = \frac{(\xi^2 - 1)\sqrt{\nu}\,\Gamma\left(\frac{\nu-1}{2}\right)}{\xi\sqrt{\pi}\,\Gamma\left(\frac{\nu}{2}\right)}; \quad b = \sqrt{\frac{\nu(\xi^4 - \xi^2 + 1)}{\xi(\nu - 2)} - a^2} \tag{12}$$

and $g^*(w;\nu)$ stands for the usual (i.e. non-standardized) Student t density with ν degrees of freedom.

This skewed Student $t(\nu;\xi)$ AR(1) - GARCH(1,1) model (hereafter denoted "Model 2") generalizes Model 1 by allowing the distribution of the innovations to be asymmetric. It contains Model 1 as a special case[10] for $\xi = 1$, has positively skewed innovations if $\xi > 1$ and negatively skewed innovations if $\xi < 1$. As Model 1 and whatever the value of the asymmetry parameter ξ, its underlying innovations only possess finite moments up to order $\delta < \nu$.

The cdf. $G(\varepsilon;\theta)$ — needed for computing the probability integral transform of the innovations underlying our proposed test — corresponding to the density $g(\varepsilon;\theta)$ of these two models are given, for Model 1, by

$$G(\varepsilon;\theta) = G^*(b\varepsilon;\nu) \tag{13}$$

where $\theta = \nu > 2$, $b = \sqrt{\frac{\nu}{\nu-2}}$ and $G^*(w;\nu)$ stands for the cdf. of the usual (i.e. non-standardized) Student t density with ν degrees of freedom[11], and, for Model 2, by

$$G(\varepsilon;\theta) = \frac{1}{\xi^2 + 1}\left[\begin{array}{l} I_{(a+b\varepsilon<0)}\,2G^*(\xi(a + b\varepsilon);\nu) \\ + I_{(a+b\varepsilon\geq 0)}\left((1 - \xi^2) + 2\xi^2 G^*\left(\frac{a + b\varepsilon}{\xi};\nu\right)\right)\end{array}\right] \tag{14}$$

where $\theta = (\nu;\xi)'$ with $\nu > 2$ and $\xi > 0$, a and b are as defined in (12) and $G^*(w;\nu)$ again denotes the cdf. of the usual (i.e. non-standardized) Student t density with ν degrees of freedom[12].

For evaluating the finite sample behavior of the different versions of our proposed test in these two models, we considered the following four data generating processes (DGP), which all maintained the AR(1) - GARCH(1,1) structure (1) with parameter values $\gamma_1 = 0$, $\gamma_2 = 0.1$, $\gamma_3 = 0.05$, $\gamma_4 = 0.1$ and $\gamma_5 = 0.8$ for the conditional mean and the conditional variance, but put into action different distributions for the innovations ε_t:

- DGP 1 : $\varepsilon_t \sim$ (standardized) Student $t(5)$
- DGP 2 : $\varepsilon_t \sim$ (standardized) skewed Student $t(5; 1.15)$
- DGP 3 : $\varepsilon_t \sim$ (standardized) generalized error distribution[13] (GED) with parameter equal to 1.3
- DGP 4 : $\varepsilon_t \sim$ mixture of (standardized) skewed Student $t(4; 1.6)$ and skewed Student $t(6; 0.65)$ in respective proportions 0.6 and 0.4

[10] For $\xi = 1$, the density (11) collapses to the density (2).

[11] This cdf. has no closed form, but vitually all statistical softwares provide a relevant numerical procedure to compute it.

[12] Needless to say, for $\xi = 1$, (14) collapses to (13).

[13] For details about the GED distribution, see Nelson (1991) or Bollerslev and al. (1994).

The distributions of the innovations corresponding to these four DGP are graphed in Figure 1. Note that except the distribution of DGP 3, they all possess only few …rst …nite moments.

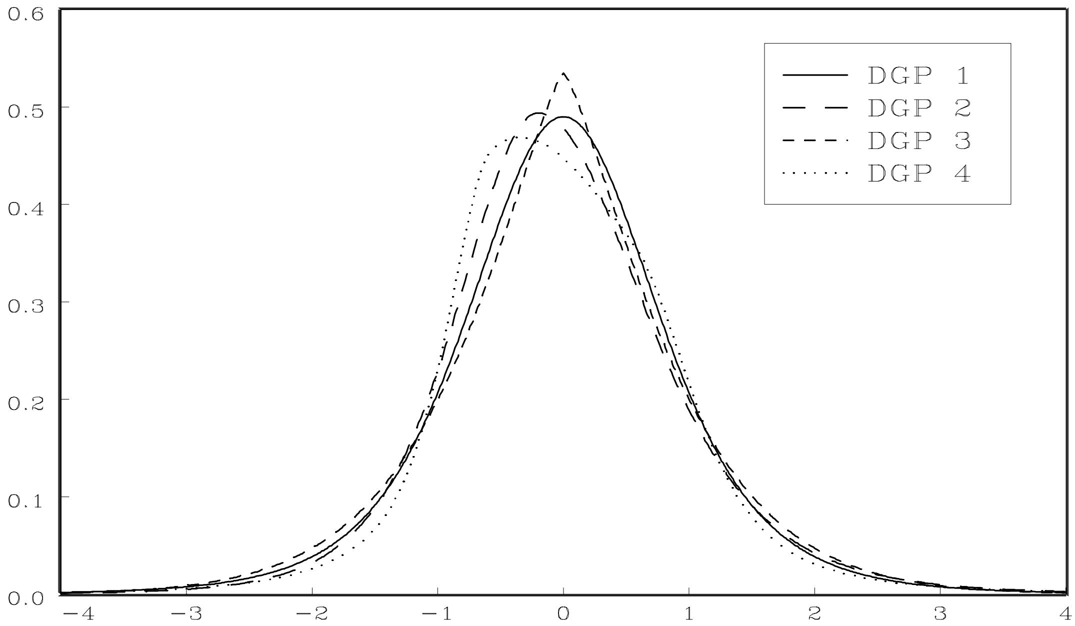

Figure 1: Distributions of the innovations

Using these four DGP, we assessed the …nite sample size and power of the different versions of our proposed test in the two models by considering, on one hand, for size evaluations, tests of Model 1 under DGP 1 and of Model 2 under DGP 1 and 2, and on the other hand, for power evaluations, tests of Model 1 under DGP 2 and 3 and of Model 2 under DGP 3 and 4.

The entire exercise was performed for three sample sizes, namely $n = 400$, $n = 800$ and $n = 1600$. In all cases, test sizes were evaluated making 5000 replications and, to alleviate computational burden, test powers making 2000 replications[14].

Table 1 reports the obtained results for tests performed at 5% nominal asymptotic level. The upper part of Table 1 groups together the estimated size of the different versions of our test for the di¤erent considered combinations of Model / DGP and sample sizes. Its lower part groups together their estimated power, likewise for the di¤erent considered combinations of Model / DGP and sample sizes. In the lower part of Table 1, for allowing to assess power separately from discrepancy between nominal and actual size, along with the estimated power of the tests is also reported (put below in brackets) an evaluation of their "size-corrected power", i.e. their power computed using as critical value, instead of their nominal asymptotic 5% critical value, an evaluation of their actual …nite sample 5% critical value calculated from the Monte-Carlo results under the "closest" null Model / DGP[15].

[14] The Monte-Carlo results reported in Table 1 already represent about 1 month of computations 24 hours a day on a 900 Mhz PC. Note that what actually takes times is not test statistics calculations but ML estimations of the models.

[15] i.e. from the Monte-Carlo results of tests of Model 1 under DGP 1 for tests of Model 1 under DGP 2 and 3, of tests of Model 2 under DGP 1 for tests of Model 2 under DGP 3, and …nally of tests of Model 2 under DGP 2 for tests of Model 2 under DGP 4.

Table 1 : Monte-Carlo results

Model / DGP	Test stat.	n = 400 Tested moments q = 2	q = 4	q = 6	n = 800 Tested moments q = 2	q = 4	q = 6	n = 1600 Tested moments q = 2	q = 4	q = 6
1=1	OPG	8:5	11:4	12:0	7:8	9:3	10:6	6:2	7:2	7:8
	PML	3:9	5:0	5:1	4:3	5:5	5:7	3:8	4:9	4:9
2=1	OPG	10:0	14:5	17:1	7:0	9:0	11:6	6:5	8:3	10:0
	PML	4:0	5:3	5:7	3:8	5:3	5:6	3:6	5:4	6:1
2=2	OPG	9:4	13:3	17:2	7:7	10:6	12:0	5:7	7:8	8:8
	PML	3:5	5:3	6:0	4:3	5:6	6:1	3:0	4:7	5:1
1=2	OPG	15:5	35:9	36:9	15:8	58:5	58:3	19:4	88:2	87:0
		(9.7)	(19.7)	(17.6)	(11.1)	(46.2)	(43.0)	(17.0)	(84.3)	(82.3)
	PML	4:9	24:2	23:6	6:9	51:0	51:2	9:2	86:5	84:4
		(6.3)	(24.1)	(22.9)	(8.3)	(49.4)	(48.4)	(12.1)	(86.7)	(84.7)
1=3	OPG	37:8	36:2	37:3	61:9	55:3	55:2	88:2	82:8	81:1
		(28.5)	(22.1)	(20.8)	(54.4)	(44.7)	(40.5)	(86.1)	(78.9)	(75.9)
	PML	27:3	23:6	22:0	56:3	47:8	45:4	85:6	80:6	76:9
		(31.1)	(23.4)	(21.1)	(59.2)	(46.0)	(43.4)	(88.3)	(81.0)	(77.3)
2=3	OPG	39:9	44:3	45:7	60:9	59:1	57:1	87:6	84:5	82:2
		(28.1)	(23.4)	(22.9)	(56.0)	(46.8)	(40.9)	(85.4)	(77.6)	(73.0)
	PML	26:6	27:0	25:2	54:5	50:2	45:1	85:6	81:8	78:4
		(30.6)	(25.9)	(23.2)	(57.8)	(49.1)	(43.2)	(87.9)	(81.0)	(75.1)
2=4	OPG	30:9	41:0	39:1	60:1	67:0	62:9	93:4	95:5	93:2
		(18.6)	(19.3)	(11.9)	(51.5)	(50.7)	(39.9)	(92.5)	(92.9)	(89.1)
	PML	12:9	26:6	22:4	46:1	58:4	53:4	90:0	94:2	91:5
		(17.0)	(25.8)	(19.0)	(48.9)	(56.8)	(49.8)	(93.0)	(94.8)	(91.4)

Reported sizes and (size-corrected) powers are expressed in percentage.

Considering …rst the size properties of the tests, it appears that the tests implemented using the M_n^{OPG} and the M_n^{PML} statistics behave very di¤erently.

According to their reputation, the OPG tests tend to be quite severely oversized, at least for n moderate and q large. As a matter of fact, for $n = 400$, the estimated sizes of the OPG tests range from 8:5% to 17:2%, the smallest size bias being encountered for $q = 2$ and the largest for $q = 6$. Things gets better when considering larger sample sizes. However, for $n = 1600$, estimated sizes still range from 5:7% to 10:0% , which is still, at least in the worst cases (i.e. for the largest q), quite in excess from their nominal 5% level.

In sharp contrast, the PML tests turn out to exhibit actual sizes remarkably close to their nominal 5% asymptotic level : all together, for the di¤erent considered cases, their estimated sizes all fall into the narrow interval [3:0%; 6:1%], indicating only a small tendency to under-reject for $q = 2$ and to over-reject for $q = 6$.

These …ndings clearly suggest that, for reliable inference, unless the sample size is very large (or simply large but then q is limited to $q = 2$), the M_n^{PML} version of our proposed test should in practice always be preferred.

Turning our attention to the power properties of the tests, it may …rst be noted that if, as a natural consequence of their size distortions, the OPG tests show off — quite importantly for n small, but only slightly for n large — higher power than their corresponding PML tests, this is no longer true when considering size-corrected power of the tests: for the same true size, the PML tests actually turn out to be as powerful — and even most of the time slightly more powerful — as their corresponding OPG tests. This of course provides a further reason for favoring in practice the M_n^{PML} version of our test.

Besides, examining the power of the tests in regard to the number q of tested moment restrictions, it may be seen that setting $q = 4$ yields tests with quite "uniformly" good[16] power against all of the various envisaged forms of misspecification, i.e. departures from the null model in terms of skewness (Model 1 under DGP 2), in terms of kurtosis (Model 1 and 2 under DGP 3) as well as in terms of "general shape" (Model 2 under DGP 4).

This is not the case of tests implemented with only $q = 2$: if setting $q = 2$ appears to work well and even the best for Model 1 and 2 under DGP 3, it turns out to work very badly for Model 1 under DGP 2, indicating that considering $q = 2$ is definitely not enough for ensuring power against various alternatives.

On the other hand, setting further $q = 6$ does not seem to pay off: it indeed yields less powerful tests than the ones obtained from $q = 4$ in almost all of the considered cases.

Overall, these observations suggest that, regarding the number of moment restrictions to consider, setting $q = 4$ is the most recommendable practical choice, and that for this choice, our proposed test appears to constitute an effective tool, able to detect various forms of distributional misspecification.

6. Conclusion

This paper proposed a diagnostic m-test for checking the distributional specification of parametric conditional heteroscedasticity models for financial data. Being based on the moments of the probability integral transform of the innovations of the assumed model rather than on the moments of the innovations themselves, the proposed test is applicable regardless whether or not, as likely to happen when working with a number of popular models such as the customary Student t GARCH model, under the null of correct specification, the innovations underlying the model possess only few first finite moments, providing thereby a convenient diagnostic test of general applicability.

Monte-Carlo evidence indicates that, put into practice setting $q = 4$ and using the M_n^{PML} statistic, or for computational easiness the M_n^{OPG} statistic if the sample size is sufficiently large, our proposed test works well both in terms of size and power, supporting thus the idea that it constitutes a valuable and hopefully fruitful empirical tool for evaluating the distributional aspect of widely used parametric conditional heteroscedasticity models.

[16] Note that, as suggested by Figure 1, in all of the considered scenarios, the alternative is very close to the null model.

References

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

Bollerslev T. and Wooldridge J. (1992), "Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances", Econometric Reviews, 11 (2), 143-172.

Bollerslev T., Engle R. and Nelson B. (1994), "Arch models", Ch. 49 in Handbook of Econometrics, Vol. 4, eds. Engle R. and McFadden D., North Holland, Amsterdam.

Davidson R. and MacKinnon J.G. (1993), Estimation and Inference in Econometrics, Oxford University Press, Oxford.

Diebold F.X., Gunther T.A. and Tay A.S. (1998), "Evaluating density forecasts with applications to …nancial risk management", International Economic Review, 39, 863-883.

Fernandez C. and Steel M.F. (1998), "On Bayesian modeling of fat tails and skewness", Journal of the American Statistical Association, 93 (441), 359-371.

Giot P. and Laurent S. (2001), "Value-at-risk for long and short trading positions", CORE Discussion Paper n± 01/22, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.

Hansen B.E. (1994), "Autoregressive conditional density estimation", International Economic Review, 35, 705-730.

Jarque C.M. and Bera A.K. (1980), "E¢cient tests for normality, homoscedasticity and serial independence of regression residuals", Economics Letters, 6, 255-259.

Johnson N.L. and Kotz S. (1970), Distributions in Statistics : Continuous Univariate Distributions - 2, John Wiley and Sons, New-York.

Lambert Ph. and Laurent S. (2001), "Modelling …nancial time series using GARCH-type models and a skewed Student density", Mimeo, Université de Liège, Liège, Belgium.

Lye J.N., Martin V.L. and Teo L.E. (1998), "Parametric distributional ‡exibility and conditional variance models with application to hourly exchange rates". Working Paper n± 98/29, International Monetary Fund, Washington, USA.

Nelson D.B. (1991), "Conditional heteroscedasticity in asset returns : a new approach", Econometrica, 59 (2), 347-370.

Newey W. (1985), "Maximum likelihood speci…cation testing and conditional moment tests", Econometrica, 53 (5), 1047-1070.

Palm F.C. (1996), "GARCH models of volatility", Ch. 17 in Handbook of Statistics, Vol. 14, eds. Maddala G.S., Rao C.R. and Vinod H.D., North Holland, Amsterdam.

Paolella M.C. (1999), "Using ‡exible GARCH models with asymmetric distributions", Working paper, Institute of Statistics and Econometrics, Christian Albrechts University, Kiel, Germany.

Tauchen G. (1985), "Diagnostic testing and evaluation of maximum likelihood models", Journal of Econometrics, 30, 415-443.

White H. (1987), "Speci…cation testing in dynamic models", Ch. 1 in Advances in Econometrics - Fifth World Congress, Vol. 1, eds. Bewley T., Cambridge University Press, New-York.

White H. (1994), Estimation, Inference and Speci…cation Analysis, Cambridge University Press, Cambridge.

Wooldridge J. (1990), "A uni…ed approach to robust, regression-based speci…cation tests", Econometric Theory, 6, 17-43.

Wooldridge J. (1991a), "On the application of robust, regression-based diagnostics to models of conditional means and conditional variances", Journal of Econometrics, 47, 5-46.

Wooldridge J. (1991b), "Speci…cation testing and quasi-maximum likelihood estimation", Journal of Econometrics, 48, 29-55.

Wooldridge J. (1994), "Estimation and inference for dependent processes", Ch. 45 in Handbook of Econometrics, Vol. 4, eds. Engle R. and McFadden D., North Holland, Amsterdam.