Journal of Empirical Finance 16 (2009) 507–523

Contents lists available at ScienceDirect

Journal of Empirical Finance

journal homepage: www.elsevier.com/locate/jempfin

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

Bernard Lejeune [1]

HEC — University of Liège, Boulevard du Rectorat, 3, B33, 4000 Liège, Belgium

ARTICLE INFO

Article history:
Received 15 October 2008
Accepted 22 November 2008
Available online 7 December 2008

JEL classification:
C12
C22
C52

Keywords:
Parametric conditional heteroscedasticity models
Distributional specification test
m-testing

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 test performs well both in terms of size and power. An empirical example illustrates the practical usefulness of the test and some of its possible extensions are outlined.

1. Introduction

Parametric conditional heteroscedasticity models, such as the standard Student *t* GARCH model of Bollerslev (1987), are a well-established and indisputably fruitful tool for analyzing financial data. If these models are routinely estimated in empirical applications, diagnostic testing of their specification, and more particularly of their distributional specification, is however much less common in practice. To help fill 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 first idea which naturally comes in mind when thinking about a diagnostic test for checking the distributional specification of parametric conditional heteroscedasticity models is, by analogy to the popular Jarque and 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 significantly different from their (estimated) theoretical values as they follow from the assumed distribution for the innovations.

This way of looking at distributional misspecification 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 such as the customary 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 2*r*-th moment of this quantity be finite. Accordingly, the above testing strategy can only be applied provided that the distribution of the innovations underlying of the model specification possesses, under the null of correct specification, finite moments at least up to order eight. If there is obviously nothing to worry about regarding this

E-mail address: B.Lejeune@ulg.ac.be.

[1] I wish to thank Luc Bauwens, participants of the econometrics seminar at Maastricht University, as well as two anonymous referees for helpful comments on earlier versions of this paper.

0927-5398/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jempfin.2008.11.005

moment condition whenever working with distributions which always possess finite moments of all orders, this is no longer true when dealing with distributions which in practice may turn out to possess only few finite moments.

As a matter of fact, empirical studies using the standard Student $t(\nu)$ GARCH model, whose by definition underlying Student $t(\nu)$ innovations only possess finite moments up to order $\delta<\nu$, typically report estimated values between 4 and 8 for the number of degrees of freedom ν, suggesting that the underlying innovations do not usually possess enough finite 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 finite moments of all orders. These include, among others, the generalized Student t GARCH model of Bollerslev et al. (1994), the skewed Student t GARCH models of Hansen (1994), Lambert and Laurent (2001) and Giot and Laurent (2003), or the generalized Student t GARCH model as considered in Lye et al. (1998).

One way to avoid this problem is to resort to non-parametric Kolmogorov–Smirnov or Cramer–Von Mises type tests allowing for dynamic models and estimated parameters under the null. Such tests have recently been proposed by respectively Bai (2003) and Inoue (1998). A major drawback of these tests is that they are difficult to implement, and thus unlikely to be widely used in empirical practice. Another possibility would be to develop, as Andrews (1988a,b), a Pearson chi-square type test not only allowing for estimated parameters under the null but also suitable for dynamics models (Andrews (1988a,b) assumes i.i.d. observations). Such a test could presumably be obtained following the general lines of Whang and Andrews (1993).

This paper takes another road. To circumvent the existence of moments 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 goodness-of-fit evaluation procedure proposed among others by Kim et al. (1998) and Diebold et al. (1998), 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 proposed test. An empirical example illustrating the practical usefulness of the test is presented in Section 6. Finally, Section 7 concludes and outlines some extensions.

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 variable 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 $\mathcal{P}$ for the conditional densities of y_t given x_t :

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

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

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

This parametric model generically describes the most commonly used class of parametric models for modelling financial data. In empirical 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 typically 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

$$\mu_t(x_t,\gamma) = \delta_0 + \delta_1 y_{t-1}, \quad h_t(x_t,\gamma) = \alpha_0 + \alpha_1 u_{t-1}^2 + \beta_1 h_{t-1} \tag{1}$$

and

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

where $\gamma=(\delta_0,\delta_1,\alpha_0,\alpha_1,\beta_1)'$, $\eta=\nu>2$, $u_{t-1}=y_{t-1}-\mu_{t-1}(x_{t-1},\gamma)$, $h_{t-1}=h_{t-1}(x_{t-1},\gamma)$, $b=\sqrt{\frac{\nu}{\nu-2}}$ 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 $\mathcal{P}$ is given by the ML estimator

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

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[4] (e.g. White (1994) or Wooldridge (1994)), if model $\mathcal{P}$ is correctly specified, i.e. if there exists some true value θ^o in Θ such that

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

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 $\mathcal{P}$, whose limiting distribution is

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

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

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

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 $\mathcal{P}$, i.e. in testing the null hypothesis that model $\mathcal{P}$ 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 be checked using robust to distributional misspecification LM-type diagnostic m-tests based on preliminary (Gaussian) pseudo-maximum likelihood estimation of the model (see White (1987, 1994), Wooldridge (1990, 1991a,b) and Bollerslev and Wooldridge (1992))[5].

[2] For a general survey on these models, see Bollerslev et 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. $g^*(w;\nu)=\frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\sqrt{\pi\nu}\Gamma\left(\frac{\nu}{2}\right)}\left(1+\frac{w^2}{\nu}\right)^{-\frac{\nu+1}{2}}$.

[4] Following Bollerslev et al. (1994), note that the verification of these regularity conditions has to be done on a case-by-base basis and is, for general models, usually difficult. The common practice in empirical studies is thus to simply proceed assuming that the necessary conditions are indeed satisfied.

[5] As outlined by a referee, such robust LM-type tests might have less power than standard LM tests when the actual distribution of the innovations is far from normality. This is the price to pay for their robustness. When this may be a real concern, i.e. when the sample size is not very large, it may be a good idea to re-check the conditional mean and conditional variance specification based on standard LM-test if no distributional specification is indeed found using your proposed test.

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

$$\hat{M}_n = \frac{1}{n}\sum_{t=1}^{n} m_t\left(y_t, x_t, \hat{\theta}_n\right), \text{ with } \quad m_t\left(y_t, x_t, \hat{\theta}_n\right) = \begin{bmatrix} \hat{e}_t^3 - \phi_3(\hat{\eta}_n) \\ \hat{e}_t^4 - \phi_4(\hat{\eta}_n) \end{bmatrix} \tag{5}$$

where

$$\hat{e}_t = e_t(y_t, x_t, \hat{\gamma}_n) = \frac{y_t - \mu_t(x_t, \hat{\gamma}_n)}{\sqrt{h_t(x_t, \hat{\gamma}_n)}} \text{ and } \quad \phi_r(\eta) = \int_{-\infty}^{+\infty} \varepsilon^r g(\varepsilon;\eta) d\varepsilon \tag{6}$$

is not significantly different from zero.

The rationale of such a testing strategy stems from the fact that under correct specification of $\mathcal{P}$, 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

$$E\left[(e_t^o)^r - \phi_r(\eta^o)\right] = 0, \; t = 1, 2, \ldots \tag{7}$$

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

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

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 Eq. (7), 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[6].

For this testing strategy to be applicable, it is necessary that, 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 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 Eq. (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[7].

As a general device to circumvent this problem while staying in the convenient m-testing framework, following the line of the graphical and informal goodness-of-fit evaluation procedure proposed among others by Kim et al. (1998) and Diebold et al. (1998), 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 is given by

$$\hat{v}_t = v_t\left(y_t, x_t, \hat{\theta}_n\right) = G(e_t(y_t, x_t, \hat{\gamma}_n); \hat{\eta}_n) = G(\hat{e}_t; \hat{\eta}_n)$$

where

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

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

According to a lemma given in Diebold et al. (1998) — who noted that the idea can be at least traced back to Rosenblatt (1952) —, it is easily seen that, whatever the precise form of the assumed density $g(\varepsilon;\eta)$, if model $\mathcal{P}$ 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 random variable over [0,1], which by definition possesses finite moments of all orders.

[6] This doesn't 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 moment restrictions corresponding to these two moments.

[7] Note in contrast that for the evoked above robust to distributional misspecification LM-type 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.

The central moments of a continuous uniform random variable over [0,1] being, according to Johnson and Kotz (1970), given by

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

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

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

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

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

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 $\mathcal{P}$ by testing through a m-test the closeness to zero of a misspecification indicator of the form

$$\hat{M}_n = \frac{1}{n}\sum_{t=1}^{n} m_t\left(y_t,x_t,\hat{\theta}_n\right), \text{ where } m_t\left(y_t,x_t,\hat{\theta}_n\right) = \begin{bmatrix} \hat{v}_t - 0.5 \\ (\hat{v}_t - 0.5)^2 - \delta(2) \\ \vdots \\ (\hat{v}_t - 0.5)^q - \delta(q) \end{bmatrix} \tag{8}$$

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. Indeed, departures from the assumed density in terms of skewness will generically result in an asymmetric distribution for $\hat{v}_t$, which may already be detected through a deviation of the first order sample moment (mean) of $\hat{v}_t$ from 0.5. Likewise, departures from the assumed density in terms of kurtosis will generically result in a distribution of $\hat{v}_t$ with a "butterfly shape" (i.e. with a hump in the middle and two wings on the sides, see Diebold et al. (1998)) or an inverted "butterfly shape", which on its side may already be detected through a deviation of the second order sample moment (variance) of $\hat{v}_t$ from $\delta(2)$ Nevertheless, for ensuring power against a large spectrum of alternatives, it is likely that setting $q = 4$ or further $q = 6$ is a better practical choice[8]. As we will see in the Monte-Carlo simulations reported below, this is indeed the case.

To conclude this discussion, remark that using the probability integral transform of the (estimated) innovations is not the only way by which a diagnostic m-test of general applicability can be obtained. Actually, using any transformation which, under correct specification, maps the innovations into a known distribution with finite moments of all orders is suitable. For example, the following composed transformation

$$\underline{\hat{v}}_t = \underline{v}_t\left(y_t,x_t,\hat{\theta}_n\right) = \Phi^{-1}(G(e_t(y_t,x_t,\hat{\gamma}_n);\hat{\eta}_n)) = \Phi^{-1}(G(\hat{e}_t;\hat{\eta}_n)) = \Phi^{-1}\left(\hat{v}_t\right)$$

where $\Phi^{-1}(\cdot)$ denotes the inverse of the standard normal (or for convenience logistic) cdf., which according to the same reasoning that above and under correct specification maps the innovations into independent of x_t and i.i.d standard normal (or logistic) random variables, could equally be used. Using this transformation, a moment-based test would likewise be obtained by checking through a m-test that the first q sample central moments of $\underline{\hat{v}}_t$ are not significantly different from those of the standard normal (or logistic) distribution.

So, why preferring the probability integral transform? First, because it is simpler. But more decisively because, as it can readily be checked by simulation, it appears that the distribution of the higher order sample central moments of uniform random variables converges much more rapidly to normality than the higher order sample central moments of standard normal (or logistic) random variables, implying that a m-test[9] based on probability integral transform should be (much) better behaved in small sample than

[8] As outlined by a referee, it is likely that the optimal choice of q is case dependent, varying with both the model and the actual data generating process. It seems however difficult to say anything general in this respect.

[9] Whose statistic is just a quadratic form in these sample central moments.

a m-test based on the above composed transformation, or composed transformation of the same kind (i.e. mapping the innovations into a Gaussian-like distribution).

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 (8) is, under general regularity conditions, given by the asymptotically chi-square statistic

$$\mathcal{M}_n = n\hat{M}_n'\hat{K}_n^{-1}\hat{M}_n \rightarrow_d \chi^2(q) \tag{9}$$

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

$$K_n^o = \frac{1}{n}\sum_{t=1}^{n} E\left[\left(m_t^o - D_n^o A_n^{o\,-1} s_t^o\right)\left(m_t^o - D_n^o A_n^{o-1} s_t^o\right)'\right] \tag{10}$$

$$= \frac{1}{n}\sum_{t=1}^{n} E\left[m_t^o m_t^{o'}\right] - \frac{1}{n}\sum_{t=1}^{n} E\left[m_t^o s_t^{o'}\right]\left(\frac{1}{n}\sum_{t=1}^{n} E\left[s_t^o s_t^{o'}\right]\right)^{-1}\frac{1}{n}\sum_{t=1}^{n} E\left[s_t^o m_t^{o'}\right] \tag{11}$$

where

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

s_t^o and H_t^o are as defined in Eq. (4), and the equality of Eqs. (10) and (11) follows from the so-called information matrix ($A_n^o = -\frac{1}{n}\sum_{t=1}^{n} E\left[s_t^o s_t^{o'}\right]$) and cross-information matrix ($D_n^o = -\frac{1}{n}\sum_{t=1}^{n} E\left[m_t^o s_t^{o'}\right]$) equalities.

The simplest operational form of Eq. (9) is obtained by taking as a consistent estimator of K_n^o the empirical counterpart of Eq. (11)

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

where

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

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

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

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

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

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 for often exhibiting poor finite sample properties[10]: 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. These poor finite sample properties follow from the fact that the covariance matrix estimator $\hat{K}_n^{\mathrm{OPG}}$ typically tends to underestimate K_n^o.

[10] See for example Davidson and MacKinnon (1993), Ch 16.

According to our experience, as a general rule, a usually better behaved in finite sample version of Eq. (9) 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)

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

constructed using as a consistent estimator[11] of K_n^o, instead of the empirical counterpart of Eq. (11), the empirical counterpart of Eq. (10)

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

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

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

The usually better finite sample properties of this alternative test statistic follow from the fact that $\hat{K}_n^{\mathrm{PML}}$ generally provides a more accurate estimator of K_n^o than the outer-product-gradient estimator $\hat{K}_n^{\mathrm{OPG}}$.

Interestingly, this alternative statistic $\mathcal{M}_n^{\mathrm{PML}}$ 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[12] of the matrix of derivatives $\hat{D}_n$ and $\hat{A}_n$, while still convenient, it is not as computationally easy as $\mathcal{M}_n^{\mathrm{OPG}}$.

Before looking at the actual finite sample performance of different possible versions of our proposed test, let us conclude this section by making a few remarks regarding the interpretation of the test. First note that the null hypothesis of the test assumes correct specification of model $\mathcal{P}$ in its entirety, i.e. not only correct distributional specification but also correct conditional mean and conditional variance specification. When rejecting the null hypothesis, it may thus be due to distributional specification, but also to conditional mean and/or conditional variance misspecification with actually no distributional misspecification. For being able to attribute a rejection of the null hypothesis to distributional misspecification, we need to be reasonably confident that the conditional mean and the conditional variance are not misspecified. As we already outlined, this may appropriately be checked using robust to distributional misspecification LM-type diagnostic *m*-tests based on preliminary (Gaussian) pseudo-maximum likelihood estimation of the model. We believe that it is a good practice to first check the mean and variance specification using these robust to distributional misspecification diagnostic tests before even estimating (by standard ML) the fully parametric model based on some assumption for the distribution of the innovations. This is in the line of the sequential "bottom–up" model construction/specification testing strategy advocated by Wooldridge (1991a).

If no sign of misspecification is found in the conditional mean and variance based on preliminary robust estimation and testing of the model, then a rejection of the null of our proposed test may sensibly be attributed to distributional misspecification. When distributional misspecification is encountered, we may try to fix it. This means considering a more suitable density for the innovations of the model. To this end, we need to identify the kind of departure from the assumed density which led to the rejection of the null. We outlined above that departures from the assumed density in terms of skewness will generically result in an asymmetric distribution for $\hat{v}_t$, while departures in terms of kurtosis will generically result in a distribution of $\hat{v}_t$ with a "butterfly shape" or an inverted "butterfly shape". Some insights about the kind of departure from the null may thus be obtained by simply examining an histogram of $\hat{v}_t$, as suggested by Diebold et al. (1998) for evaluating density forecasts. According to our experience, although somewhat less convenient, a more fruitful approach is however to graphically compare the ML estimated null density $g(\varepsilon;\hat{\eta}_n)$ and some non-parametric estimate of the density of the (Gaussian pseudo-maximum likelihood) estimated innovations $\hat{e}_t$, as we illustrate in the empirical example reported below.

A last remark: a maintained hypothesis of our parametric model $\mathcal{P}$ is that the innovations ε_t are iid. and independent of x_t, an assumption which rules out dynamics in the higher-order moments of ε_t, such as considered in the Hansen's (1994) generalized GARCH-type model where the shape parameters of the postulated innovations density are allowed to vary as a parametric function

[11] Note that both $\hat{K}_n^{\mathrm{OPG}}$ and $\hat{K}_n^{\mathrm{PML}}$ 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 $\hat{K}_n = \frac{1}{n}\sum_{t=1}^{n}\hat{m}_t\hat{m}_t' - \hat{D}_n\left(\frac{1}{n}\sum_{t=1}^{n}\hat{s}_t\hat{s}_t'\right)^{-1}\hat{D}_n'$, a characteristic which a priori discards them as relevant candidates for constructing an operational version of Eq. (9).

[12] 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 $\mathcal{M}_n^{\mathrm{OPG}}$ and $\mathcal{M}_n^{\mathrm{PML}}$ forms of our proposed test may be obtained upon request from the author.

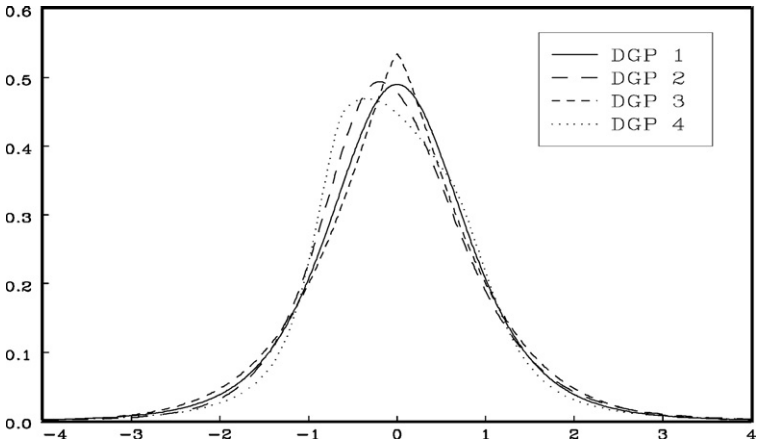

Fig. 1. Distributions of the innovations.

of x_t. Although routinely assumed, this maintained hypothesis might in practice not hold. Our proposed test, which concentrates on detecting departures from the shape of the assumed innovations density, is unlikely to have much power against such dynamic distributional misspecification. This possibility has however to be kept in mind, both when the null is rejected and when it is not. We will come back to this point and suggest a possible way to check for such dynamic distributional misspecification in our concluding comments, where some extensions of our proposed test are outlined.

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 $\mathcal{M}_n^{\mathrm{OPG}}$ and $\mathcal{M}_n^{\mathrm{PML}}$ 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 Steel (1998)[13], whose density $g(\varepsilon;\eta)$ may be written

$$g(\varepsilon;\eta) = \frac{2b\kappa}{\kappa^2+1}\left[\mathbb{I}_{(a+b\varepsilon<0)}g^*(\kappa(a+b\varepsilon);\nu) + \mathbb{I}_{(a+b\varepsilon\geq 0)}g^*\left(\frac{a+b\varepsilon}{\kappa};\nu\right)\right] \tag{12}$$

where $\eta=(\nu,\kappa)'$ with $\nu>2$ and $\kappa>0$, while $\mathbb{I}_{(.)}$ is a 0–1 indicator function,

$$a = \frac{\left(\kappa^2-1\right)\sqrt{\nu}\Gamma\left(\frac{\nu-1}{2}\right)}{\kappa\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)}, \quad b = \sqrt{\frac{\nu(\kappa^4-\kappa^2+1)}{\kappa(\nu-2)} - a^2} \tag{13}$$

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

This skewed Student $t(\nu,\kappa)$ 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 when $\kappa=1$, has positively skewed innovations if $\kappa>1$ and negatively skewed innovations if $\kappa<1$. As Model 1 and whatever the value of the asymmetry parameter κ, its underlying innovations only possess finite moments up to order $\delta<\nu$.

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

$$G(\varepsilon;\eta) = G^*(b\varepsilon;\nu) \tag{14}$$

[13] The use of this skewed Student $t(\nu,\kappa)$ distribution in the context of parametric GARCH models has been considered and advocated in Lambert and Laurent (2001) and Giot and Laurent (2003).

Table 1
Monte-Carlo results.

Model/ DGP	Test stat.	$n=400$			$n=800$			$n=1600$		
		Tested moments			Tested moments			Tested moments		
		$q=2$	$q=4$	$q=6$	$q=2$	$q=4$	$q=6$	$q=2$	$q=4$	$q=6$
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.

where $\eta=\nu>2$, $b=\sqrt{\frac{\nu}{\nu-2}}$ and $G^*(w;\nu)$ stands for the cdf. of the usual (i.e. non-standardized) Student t density with ν degrees of freedom[14], and, for Model 2, by

$$G(\varepsilon;\eta)=\frac{1}{\kappa^2+1}\left[\begin{array}{l}\mathbb{I}_{(a+b\varepsilon<0)}2G^*(\kappa(a+b\varepsilon);\nu)\\ +\mathbb{I}_{(a+b\varepsilon\geq 0)}\left(\left(1-\kappa^2\right)+2\kappa^2G^*\left(\frac{a+b\varepsilon}{\kappa};\nu\right)\right)\end{array}\right] \tag{15}$$

where $\eta=(\nu,\kappa)'$ with $\nu>2$ and $\kappa>0$, a and b are as defined in Eq. (13) and $G^*(w;\nu)$ again denotes the cdf. of the usual (i.e. non-standardized) Student t density with ν degrees of freedom.

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 $\delta_0=0$, $\delta_1=0.1$, $\alpha_0=0.05$, $\alpha_1=0.1$ and $\beta_1=0.8$ for the conditional mean and the conditional variance, but put into action different distributions for the innovations ε_t :

- DGP 1 : ε_t~(standardized) Student $t(5)$
- DGP 2 : ε_t~(standardized) skewed Student $t(5, 1.15)$
- DGP 3 : ε_t~(standardized) generalized error distribution[15] (GED) with parameter equal to 1.3
- DGP 4 : ε_t~mixture of (standardized) skewed Student $t(4, 1.6)$ and skewed Student $t(6, 0.65)$ in respective proportions 0.6 and 0.4

The distributions of the innovations corresponding to these four DGP are graphed in Fig. 1. Note that except the distribution of DGP 3, they all possess only few first finite moments.

Using these four DGP, we assessed the finite 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.

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 different considered combinations of Model/DGP and sample sizes. Its lower part groups together their estimated power, likewise for the different 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

[14] This cdf. has no closed form, but virtually all statistical software provide a relevant numerical procedure to compute it.
[15] For details about the GED distribution, see Nelson (1991) or Bollerslev et al. (1994).

Table 2
Additional Monte-Carlo results.

Model /DGP	Test stat.	$n=400$			$n=800$			$n=1600$		
		Test type			Test type			Test type		
		PIT 4	JB*	KS	PIT 4	JB*	KS	PIT 4	JB*	KS
1/1	OPG	11.4	82.7	12.4	9.3	81.0	14.4	7.2	79.0	14.6
	PML	5.0	3.4		5.5	2.0		4.9	1.7	
1/lb	OPG	13.9	53.2	12.3	10.4	46.5	13.6	8.7	38.5	l4.7
	PML	5.3	2.7		5.9	1.7		5.8	2.1	
1/2	OPG	35.9	86.3	17.8	58.5	88.6	22.2	88.2	89.8	44.2
		(19.7)	(5.3)	(5.9)	(46.2)	(6.2)	(5.1)	(84.3)	(8.6)	(7.5)
	PML	24.2	23.6		51.0	34.4		86.5	51.5	
		(24.1)	(28.2)		(49.4)	(47.6)		(86.7)	(67.1)	
1/3	OPG	36.2	92.3	8.6	55.3	96.2	13.7	82.8	98.3	25.3
		(22.1)	(6.0)	(1.2)	(44.7)	(5.9)	(1.1)	(78.9)	(8.4)	(2.9)
	PML	23.6	2.3		47.8	2.1		80.6	3.5	
		(23.4)	(3.9)		(46.0)	(6.5)		(81.0)	(12.6)	
1/2*b*	OPG	35.0	74.3	17.2	52.4	76.3	21.7	83.9	89.3	38.9
		(14.9)	(10.7)	(6.2)	(37.5)	(17.8)	(6.3)	(76.0)	(33.5)	(7.1)
	PML	21.0	25.0		44.6	43.4		81.0	79.2	
		(20.0)	(34.6)		(41.8)	(59.1)		(79.0)	(85.1)	
1/3*b*	OPG	29.9	84.0	6.6	44.0	89.4	9.0	65.3	94.3	17.3
		(15.4)	(27.6)	(1.3)	(31.3)	(43.7)	(1.4)	(56.0)	(63.4)	(1.3)
	PML	18.5	2.4		36.6	2.8		61.1	3.7	
		(17.8)	(4.6)		(34.4)	(6.4)		(59.0)	(9.7)	

Reported sizes and (size-corrected) powers are expressed in percentage.
*Rejection percentages among the cases where the statistic may actually be computed.

Table 3
Percentage of cases where the JB type test cannot be computed.

Model/DGP	$n=400$	$n=800$	$n=1600$
1/1	12.5	6.4	1.9
1/1b	0.1	0.0	0.0
1/2	15.5	8.7	3.0
1/3	4.8	1.2	0.0
1/2b	0.1	0.0	0.0
1/3b	0.1	0.0	0.0

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 finite sample 5% critical value calculated from the Monte-Carlo results under the "closest" null Model/DGP[16].

Considering first the size properties of the tests, it appears that the tests implemented using the $\mathcal{M}_n^{\mathrm{OPG}}$ and the $\mathcal{M}_n^{\mathrm{PML}}$ statistics behave very differently.

According to their reputation, the OPG tests tend to be quite severely over-sized, at least for moderate n and large q. 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 different 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 findings 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 $\mathcal{M}_n^{\mathrm{PML}}$ version of our proposed test should in practice always be preferred.

Turning our attention to the power properties of the tests, it may first 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 $\mathcal{M}_n^{\mathrm{PML}}$ 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[17] power against all of the various envisaged forms of misspecification, i.e. departures

[16] 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 finally of tests of Model 2 under DGP 2 for tests of Model 2 under DGP 4.

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

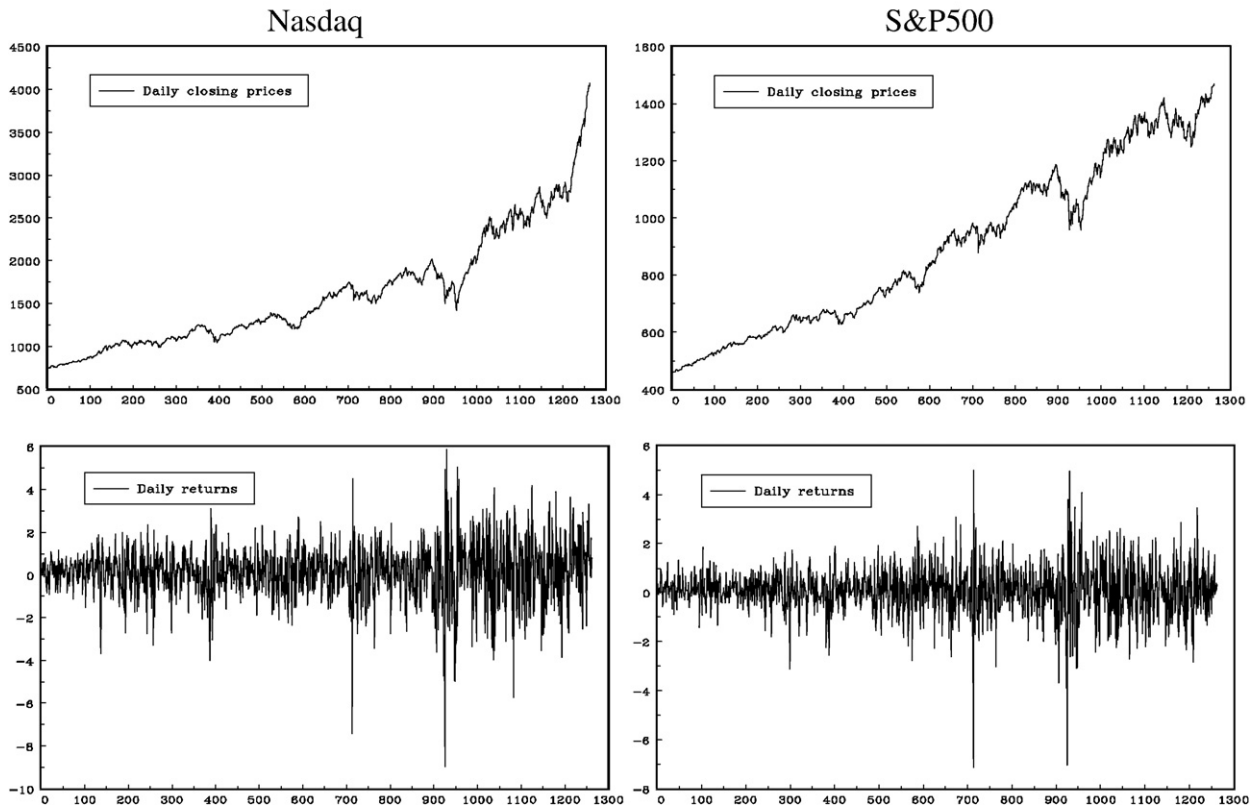

Fig. 2. Daily closing prices and returns of the Nasdaq and S&P500 stock indexes from January 3,1995 to December 31,1999.

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. This however doesn't mean that setting $q=6$ is irrelevant : it just appears to be useless for the precise forms of alternative considered here. As a matter of fact, the empirical illustration below demonstrates that setting $q=6$ may in some cases be required for being able to detect misspecification.

Overall, these observations suggest that, regarding the number of moment restrictions to consider, setting $q=4$ or for safety $q=6$ 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.

Judiciously implemented, our proposed test appears to work well. But how does it compare to other possible tests? To get insights about this question, we performed some additional simulation experiments. In these experiments, we compared the finite sample performance of our proposed test to the one of two alternative tests : on one hand, the initially envisioned Jarque-Bera like m-test based on the misspecification indicator (5), which is only valid if the innovations possess enough finite moments, and on the other hand, the Kolmogorov–Smirnov type test recently proposed by Bai (2003). This test allows for dynamic models and estimated parameters under the null, and as our proposed test resorts to probability integral transform, but checks its accordance with a uniform distribution through a Kolmogorov–Smirnov type statistic rather than through moments.

More precisely, we considered checking the distributional specification of the customary Student $t(\nu)$ AR(1)–GARCH(1,1) Model 1 using the $\mathcal{M}_n^{\mathrm{OPG}}$ and $\mathcal{M}_n^{\mathrm{PML}}$ versions of our test with $q=4$, and likewise using the $\mathcal{M}_n^{\mathrm{OPG}}$ and $\mathcal{M}_n^{\mathrm{PML}}$ forms of the Jarque–Bera like m-test based on the misspecification indicator (5)[18]. The Kolmogorov–Smirnov type test, which involves integration, was computed as suggested in Appendix B of the Bai's (2003) paper[19].

[18] Provided that the innovations possess enough finite moments, the general chi-square statistic $\mathcal{M}_n = n\hat{M}_n' \hat{K}_n^{-1} \hat{M}_n$ is again the proper m-test statistic for checking the closeness to zero of the misspecification indicator (5). Note that for Model 1, which assumes standardized Student $t(\eta)$ innovations, $\phi_3(\eta)=0$ and $\phi_4(\eta) = 3 + \frac{6}{\eta-4}(\eta>4)$.

[19] Note that, in the present model where the null distribution of the innovations depends on parameters which have to be estimated, the $g(r)$ vector needed for computing the test statistic is, in the notation of Bai (2003), given by $g(r) = (g_1, g_2, g_3, g_4)' = \left(r, f(F^{-1}(r)), f(F^{-1}(r))F^{-1}(r), \frac{\partial F(F^{-1}(r))}{\partial \lambda}\right)'$.

Table 4
Gaussian pseudo-maximum likelihood estimates.

	Nasdaq		S&P500	
Parameter	Estimate	Std. err.	Estimate	Std. err.
δ_0	0.0953	0.0298	0.0774	0.0256
δ_1	0.1011	0.0306	0.0563	0.0287
δ_3	–	–	−0.0442	0.0319
δ_5	–	–	−0.0546	0.0310
α_0	0.0570	0.0288	0.0197	0.0111
α_1	0.1539	0.0380	0.0783	0.0240
β_1	0.8333	0.0504	0.9207	0.0267
d	0.4196	0.1077	0.8308	0.1695
p	1.1412	0.2710	0.9711	0.2623

Reported standard errors are robust PML standard errors.

Table 5
Diagnostic m-tests of the conditional mean and variance.

	Nasdaq		S&P500	
Lags	Statistic	P-value	Statistic	P-value
Cond. mean tests				
$\tau=1$	0.0034	0.9533	0.3841	0.5354
$\tau=2$	0.2750	0.8715	0.4148	0.8112
$\tau=5$	2.0074	0.8481	3.9237	0.5605
$\tau=10$	3.8134	0.9554	13.2739	0.2088
Cond. variance tests				
$\tau=1$	0.6981	0.4034	1.4465	0.2291
$\tau=2$	0.7104	0.7010	2.5878	0.2742
$\tau=5$	2.4529	0.7836	3.9338	0.5590
$\tau=10$	5.0564	0.8874	6.9133	0.7336

For assessing the finite sample size and power of these tests, we used the same AR(1)–GARCH(1,1) data generating processes DGP 1, DGP 2 and DGP 3 than above, as well as three variants of them for the distribution of the innovations ε_t:

- DGP 1b: $\varepsilon_t \sim$ (standardized) Student $t(9)$
- DGP 2b: $\varepsilon_t \sim$ (standardized) skewed Student $t(9, 1.15)$
- DGP 3b: $\varepsilon_t \sim$ (standardized) generalized error distribution (GED) with parameter equal to 1.4

DGP 1b, DGP 2b and DGP 3b are basically the same as DGP 1, DGP 2 and DGP3, except that they exhibit less thick tails.

For size evaluations, we considered tests of Model 1 under DGP 1 and 1b. Note that the Jarque–Bera like m-test of Model 1 is not valid under DGP 1 (ε_t possess less than five finite moments), but is valid under DGP 1b (as required, ε_t possess first eight finite moments). This will allow us to compare the behavior of the test when it is and when it is not theoretically valid. For power evaluations, we considered tests of Model 1 under DGP 2, 3, 2b and 3b. As above, the exercise was performed for $n=400$, $n=800$ and $n=1600$, test sizes were evaluated making 5000 replications and test powers making 2000 replications.

Table 2 reports the obtained results for tests performed at 5% nominal asymptotic level. In this table, the test types PIT 4, JB and KS denote respectively our proposed test with $q=4$, the Jarque–Bera like m-test and the Bai's (2003) Kolmogorov–Smirnov type test. As Table 1, the upper part of Table 2 groups together the estimated test sizes, and its lower part groups together the estimated test powers as well as (put below in brackets) the estimated test "size-corrected powers" computed using as critical value, instead of their nominal asymptotic 5% critical value, an evaluation of their actual finite sample 5% critical value calculated from the Monte-Carlo results under the "closest" null Model/DGP[20].

Before looking at the results reported in Table 2, note that the Jarque–Bera like m-test is in practice not always computable : whenever the estimated number of degrees of freedom $\hat{\eta}_n$ of the (standardized) Student innovations of Model 1 is lower or equal to 4, because the theoretical fourth order moment $\phi_4(\eta) = 3 + \frac{6}{\eta-4}$ only exists for $\eta>4$, the test statistic can simply not be computed. Table 3 reports the percentage of cases where this happens for the different considered DGP and sample sizes.

As it could be expected, this primarily happens when the sample size is small and the DGP is (DPG 1) or is similar to (DGP 2 and 3) a (standardized) Student with only a few degrees of freedom, i.e. primarily when the test is actually not valid (Model 1 under DGP 1). But

[20] 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, and of tests of Model 1 under DGP 1b for tests of Model 1 under DGP 2b and 3b.

Table 6
Maximum likelihood estimates under Student $t(\nu)$ and skewed Student $t(\nu, \kappa)$ distributional assumptions.

	Nasdaq				S&P500			
	Student		Skewed Student		Student		Skewed Student	
Parameter	Estimate	Std. err.	Estimate	Std. err.	Estimate	Std. err.	Estimate	Std. err.
δ_0	0.1258	0.0312	0.0934	0.0303	0.1050	0.0238	0.0908	0.0242
δ_1	0.0982	0.0299	0.0892	0.0306	0.0342	0.0284	0.0287	0.0286
δ_3	–	–	–	–	−0.0663	0.0290	−0.0733	0.0285
δ_5	–	–	–	–	−0.0481	0.0296	−0.0518	0.0295
α_0	0.0408	0.0242	0.0361	0.0222	0.0142	0.0071	0.0150	0.0068
α_1	0.1285	0.0351	0.1286	0.0352	0.0713	0.0165	0.0696	0.0162
β_1	0.8637	0.0473	0.8681	0.0451	0.9269	0.0190	0.9273	0.0182
d	0.4267	0.1119	0.3911	0.1065	0.7515	0.1848	0.7636	0.1896
p	1.2046	0.2869	1.3059	0.2829	1.1223	0.2841	1.1509	0.2980
ν	10.396	2.7976	11.162	3.1681	7.6476	1.6751	8.6136	2.0546
κ	–	–	0.7989	0.0310	–	–	0.8881	0.0387

Reported standard errors are usual ML standard errors.

remark that it may also happen when the test is valid (Model 1 under DPG 1b). As a result of this problem, the JB test sizes and powers reported in Table 2 are the rejection percentages of the test only among the cases where the test statistic may actually be computed.

Returning to Table 2 and looking for a start at the size properties of the different tests, it may first be noted that what we already outlined for our PIT 4 test of Model 1 under DPG 1 still applies to the PIT 4 test of Model 1 under DPG 1b: if its $\mathcal{M}_n^{\mathrm{OPG}}$ version tends to be, at least for moderate n, quite severely over-sized, its $\mathcal{M}_n^{\mathrm{PML}}$ version exhibits actual sizes remarkably close to their nominal 5% asymptotic level.

Considering next the JB test, it turns out that its $\mathcal{M}_n^{\mathrm{OPG}}$ version is spectacularly over-sized. For $n=400$, its estimated size is 82.7% for Model 1 under DGP 1, i.e. when the test is actually not valid, and still 53.2% for Model 1 under DGP 1b, i.e. when the test is valid. For $n=1600$, things do not really improve when the test is not valid and only slightly improve — the estimated size is still 38.5% — when it is valid. Needless to say, given these finite sample sizes, the $\mathcal{M}_n^{\mathrm{OPG}}$ version of the JB test should never be used in practice.

In contrast, the $\mathcal{M}_n^{\mathrm{PML}}$ version of the JB test appears to be better behaved. Its estimated size ranges from 1.7% to 3.4%, indicating a systematic tendency to under-reject the null, which may be acceptable in practice if it does not affect the power of the test too much. Surprisingly, there is no sharp difference between the case where the test is valid (Model 1 under DGP 1b) and the case where the test is not valid (Model 1 under DGP 1). However, for safety, it is certainly not recommendable to widely use the $\mathcal{M}_n^{\mathrm{PML}}$ version of the JB test when its validity is questionable. Further, recall that when it is not valid, the test statistic may simply not be computable.

Looking finally at the KS test, it appears that, as the $\mathcal{M}_n^{\mathrm{OPG}}$ version of our PIT 4 test for moderate n, it is quite severely over-sized (its estimated size ranges from 12.3% to 14.7%), and that even for $n=1600$.

To summarize, for reliable inference, the only reasonable alternative to the $\mathcal{M}_n^{\mathrm{PML}}$ version (or the $\mathcal{M}_n^{\mathrm{OPG}}$ version if n is very large) of our proposed test is the $\mathcal{M}_n^{\mathrm{PML}}$ version of the JB test, at least when its validity is not questionable. Both the $\mathcal{M}_n^{\mathrm{OPG}}$ version of the JB test and the KS test indeed appear too much over-sized to be used in practice.

Turning our attention to the power properties of the tests, again what we already outlined for our PIT 4 test of Model 1 under DPG 2 and 3 still applies to the PIT 4 test of Model 1 under DPG 2b and 3b : our PIT 4 test appears to have quite uniformly good power against the different envisaged forms of misspecification and, considering size-corrected power, its $\mathcal{M}_n^{\mathrm{PML}}$ version turns out to be as powerful — and even slightly more powerful — as its $\mathcal{M}_n^{\mathrm{OPG}}$ version.

In sharp contrast, despite its quite severe finite sample size bias, the KS test exhibits rather low power against the different envisaged forms of misspecification. This of course provides a further argument for avoiding to use it in practice.

As suggested by its size-corrected power, the high power of the $\mathcal{M}_n^{\mathrm{OPG}}$ version of the JB test prominently stems from its huge finite sample size bias, which makes it unusable. Because it is actually under-sized, this is not the case of its $\mathcal{M}_n^{\mathrm{PML}}$ version. It however appears that if the $\mathcal{M}_n^{\mathrm{PML}}$ version of the JB test has roughly comparable power to the $\mathcal{M}_n^{\mathrm{PML}}$ version of our PIT 4 test for Model 1 under DGP 2 and 2b, it has much lower power for Model 1 under DGP 3 and 3b, indicating that checking only the third and fourth order moments of the innovations may actually not be enough for ensuring power against various alternatives. Note that adding higher order moments to the misspecification indicator (5) would probably help, but it would make the existence of moments problem which motivated our proposed test even more serious.

Overall, considering both size and power, it finally appears from Table 2 that, judiciously implemented, our proposed test not only works well, but compares favorably with both the initially envisioned Jarque–Bera like m-test and the Bai's (2003) Kolmogorov–Smirnov type test.

6. Empirical illustration

To illustrate the practical usefulness of our testing procedure, we hereafter present the results of the estimation and (distributional) specification testing of parametric conditional heteroscedasticity models for the Nasdaq and S&P500 daily returns over the period January 3, 1995 to December 31, 1999 (1263 observations). Both series[21] (closing prices p_t and returns defined as $y_t = 100\times(\ln p_t - \ln p_{t-1})$) are graphed in Fig. 2.

[21] I would like to thank Sébastien Laurent for kindly providing me the data.

Table 7
Diagnostic distributional *m*-tests.

	Nasdaq		S&P500	
Tested moments	Statistic	*P*-value	Statistic	*P*-value
Normality tests				
$q=4$	34.116	0.0000	21.446	0.0003
$q=6$	34.176	0.0000	40.084	0.0000
Student tests				
$q=4$	27.034	0.0000	6.892	0.1417
$q=6$	28.553	0.0001	24.125	0.0005
Skewed Student tests				
$q=4$	5.132	0.2740	27.455	0.0000
$q=6$	7.084	0.3132	30.204	0.0000

For both series, as a starting point, we considered an AR(ρ)–APARCH(1,1) model[22], which means setting for the conditional mean $\mu_t(\cdot)$ and the conditional variance $h_t(\cdot)$ of the returns y_t:

$$\mu_t(x_t,\gamma)=\delta_0+\sum_{j=1}^{\rho}\delta_j y_{t-j} \quad \text{and} \quad h_t(x_t,\gamma)=\left[\alpha_0+\alpha_1(|u_{t-1}|-du_{t-1})^p+\beta_1 h_{t-1}^{\frac{p}{2}}\right]^{\frac{2}{p}}$$

We first estimated, for each series, the parameters of this model by Gaussian pseudo-maximum likelihood[23] (i.e. without making any distributional assumption). For the Nasdaq series, the dynamics of the conditional mean at first sight appeared to be adequately captured by a simple AR(1), while it turned out that additional AR(3) and AR(5) terms were apparently required for the S&P500 series. Table 4 reports the parameter estimates obtained for each series.

Note that regarding the conditional variance, both series exhibit a significant leverage effect ($d>0$) and a "power parameter" p significantly different from 2, which means that the standard GARCH specification is rejected in the two series.

We then checked, for each series, the correctness of the conditional mean and conditional variance specification of the postulated model by testing through *m*-tests the closeness to zero of, for the conditional mean, misspecification indicators of the form

$$\hat{M}_n=\frac{1}{n}\sum_{t=1}^{n} m_t(y_t,x_t,\hat{\gamma}_n), \quad \text{where} \quad m_t(y_t,x_t,\hat{\gamma}_n)=\begin{bmatrix}\hat{e}_t\hat{e}_{t-1}\\ \hat{e}_t\hat{e}_{t-2}\\ \vdots\\ \hat{e}_t\hat{e}_{t-\tau}\end{bmatrix}$$

and for the conditional variance, misspecification indicators of the form

$$\hat{M}_n=\frac{1}{n}\sum_{t=1}^{n} m_t(y_t,x_t,\hat{\gamma}_n), \quad \text{with} \quad m_t(y_t,x_t,\hat{\gamma}_n)=\begin{bmatrix}\left(\hat{e}_t^2-1\right)\left(\hat{e}_{t-1}^2-1\right)\\ \left(\hat{e}_t^2-1\right)\left(\hat{e}_{t-2}^2-1\right)\\ \vdots\\ \left(\hat{e}_t^2-1\right)\left(\hat{e}_{t-\tau}^2-1\right)\end{bmatrix}$$

where $\hat{e}_t$ is as defined in Eq. (6) and for lags $\tau=1, 2, 5, 10$.

These *m*-tests, which basically look at whether or not it remains some autocorrelations (up to order τ) in the estimated innovations and squared innovations of the model[24], were computed using the same PML statistic[25] that $\mathcal{M}_n^{\mathrm{PML}}$, whose validity, as desirable in this preliminary step, doesn't rely on any distributional assumption. The results of the tests are reported in Table 5.

As it can be seen from this table, for each series, no sign of patent misspecification was found, neither in the conditional mean nor in the conditional variance.

[22] On the APARCH (Asymmetric Power ARCH) specification for the conditional variance, see Ding et al. (1993). Note that this flexible form nests at least seven GARCH-type specifications.

[23] See Gourieroux et al. (1984), Gourieroux and Monfort (1995) or Bollerslev and Wooldridge (1992). All estimates, as well as subsequent specification tests, have been computed using Gauss.

[24] Quite obviously, the statistical rationale of these tests stems from the fact that under conditional mean and conditional variance correct specification, the Gaussian PML estimator $\hat{\gamma}_n$ is consistent for γ^o and for any τ, $E[e_t^o e_{t-\tau}^o]=0$ and $E[(e_t^{o2}-1)\,(e_{t-\tau}^{o2}-1)]=0$, $t=1,2,\ldots$, where $e_t^o=e_t(y_t,x_t,\gamma^o)$, while under misspecification, $\hat{\gamma}_n$ is consistent for some pseudo-true value γ_n^* (generally) different from γ^o and (again generally) $E[e_t^* e_{t-\tau}^*]\neq 0$ and $E[(e_t^{*2}-1)\,(e_{t-\tau}^{*2}-1)]\neq 0$, $t=1,2,\ldots$, where $e_t^*=e_t(y_t,x_t,\gamma_n^*)$.

[25] It is identical to $\mathcal{M}_n^{\mathrm{PML}}$, with l_t and s_t denoting respectively the Gaussian pseudo log-likelihood of observations t and its first derivatives, and where the only involved parameters are the vector γ of mean and variance parameters.

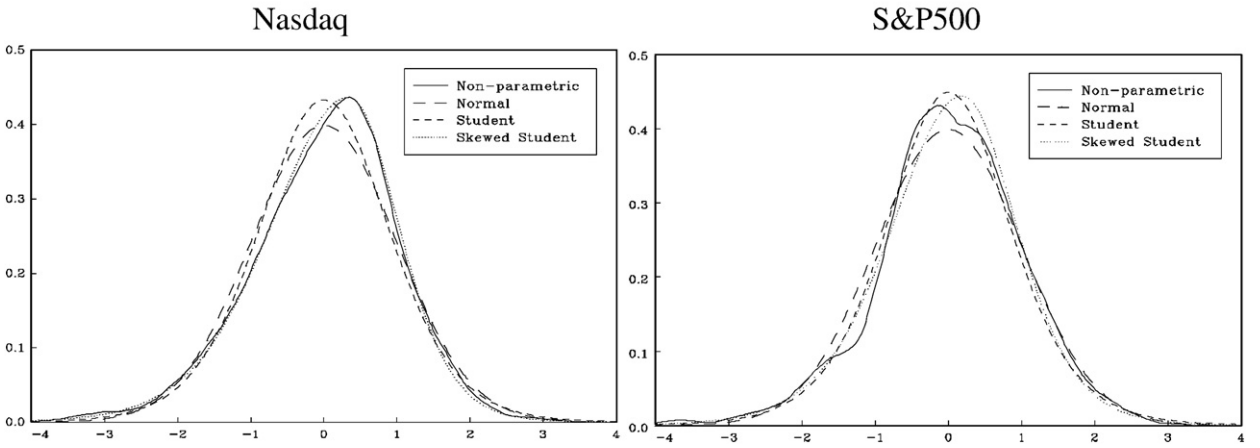

Fig. 3. Parametric and non-parametric estimates of the distribution of the innovations.

Taking the correctness of the conditional mean and variance for granted, we finally turned our attention to the distributional issue. At this stage, the question was : may the density of the innovations of the series be adequately modelled by some usual distribution?

We considered three possible candidates : a (standard) normal distribution, a (standardized) Student $t(\nu)$ distribution and the (standardized) skewed Student $t(\nu,\kappa)$ distribution briefly described above. For each of these distributional assumptions and each series, the parameters of the assumed model were estimated by maximum likelihood and then its distributional specification checked through our proposed diagnostic test implemented using the $\mathcal{M}_n^{\mathrm{PML}}$ statistic[26] and for $q=4$ and $q=6$.

The ML parameter estimates of the model under the Student $t(\nu)$ and skewed Student $t(\nu,\kappa)$ distributional assumptions are for both series reported in Table 6. Of course, the parameter estimates of the model under normality are simply those reported in Table 4.

Some variations across the conditional mean and conditional variance parameter estimates obtained under the different distributional assumptions may be observed, but they remain moderate for both series. On the other hand, the estimated distributional parameters of the most flexible skewed Student $t(\nu,\kappa)$ model suggests that the actual distribution of the innovations is both fat-tailed and negatively skewed, and that for each series[27].

As a matter of fact, the results of our distributional m-tests reported in Table 7 show that the (standard) normal and the (standardized) Student $t(\nu)$ distributional assumptions are both (strongly) rejected in each series. Note that in the case of the S&P500 series, the Student assumption is only clearly rejected when performing the test with $q=6$.

However, Table 7 also reveals that if the skewed Student $t(\nu,\kappa)$ distributional assumption seems to be adequate for the Nasdaq series, it is not the case for the S&P500 series. Fig. 3 graphically illustrates this result.

In Fig. 3 are represented, for each series, along with the standard normal distribution, on one hand, the ML estimated (standardized) Student $t(\hat{\nu})$ and (standardized) skewed Student $t(\hat{\nu},\hat{\kappa})$ distributions of the innovations, and on the other hand, a non-parametric kernel estimate of the distribution of the innovations based on the estimated innovations obtained from the Gaussian PML estimation of the model. It can be seen that if for the Nasdaq series there is a close accordance between the non-parametric estimate and the ML estimated (standardized) skewed Student $t(\hat{\nu},\hat{\kappa})$ distribution, it is far from being the case for the S&P500 series. For this series, according to the non-parametric estimate, it may be expected that finding an adequate parametric distribution for the innovations would actually be a difficult task.

7. Conclusion and extensions

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.

[26] Note that for the normality tests, since the standard normal distribution has no shape parameter, the only parameters involved in the $\mathcal{M}_n^{\mathrm{PML}}$ statistic are the vector γ of mean and variance parameters.

[27] By the way, note that if the estimated numbers of degree of freedom of the student and skewed student distributions are in three out of four cases higher than 8, the confidence intervals for these estimated numbers of degree of freedom don't rule out that, under the assumption of correct distributional specification, the true numbers of degree of freedom are actually smaller than 8, i.e. that the distributions actually possess less than first eight order finite moments.

Monte-Carlo evidence indicated that, put into practice setting $q=4$ or for safety $q=6$ and using the $\mathcal{M}_n^{\mathrm{PML}}$ statistic, or for computational easiness the $\mathcal{M}_n^{\mathrm{OPG}}$ statistic if the sample size is sufficiently large, our proposed test works well both in terms of size and power, and compares favorably with both a Jarque–Bera like m-test and the Bai's (2003) Kolmogorov–Smirnov type test. An empirical example further illustrated the relevance of our proposed testing procedure as a fruitful tool for evaluating the distributional aspect of widely used parametric conditional heteroscedasticity models.

As already outlined, a maintained hypothesis of our generic parametric model $\mathcal{P}$ is that the innovations ε_t are iid. and independent of x_t, an assumption which rules out dynamics in the higher-order moments of ε_t, such as some form of dynamic skewness or dynamic kurtosis[28]. Although standard, this maintained hypothesis might not hold.

Our proposed test, which concentrates on detecting departures from the shape of the assumed innovations density, is unlikely to have much power against such higher-order dynamic distributional misspecification. As suggested by a referee, using our framework, the presence of such dynamic distributional misspecification could however be checked by looking at autocorrelations in the probability integral transform $\hat{v}_t$, i.e. by checking through a m-test the closeness to zero of a misspecification indicator of the form $\hat{M}_n=\frac{1}{n}\sum_{t=1}^{n} m_t\left(y_t,x_t,\hat{\theta}_n\right)$, with $m_t\left(y_t,x_t,\hat{\theta}_n\right)=\left[\hat{v}_t^{*}\hat{v}_{t-1}^{*},\ldots,\hat{v}_t^{*}\hat{v}_{t-\tau}^{*},\ldots,\left(\hat{v}_t^{*q}-\delta(q)\right)\left(\hat{v}_{t-1}^{*q}-\delta(q)\right),\ldots,\left(\hat{v}_t^{*q}-\delta(q)\right)\left(\hat{v}_{t-\tau}^{*q}-\delta(q)\right)\right]'$, where $\hat{v}_t^{*}=(\hat{v}_t-0.5)$, for some integer q and some lag τ. This m-test could properly be performed using the $\mathcal{M}_n^{\mathrm{OPG}}$ statistic or, for presumably better finite sample behavior, the $\mathcal{M}_n^{\mathrm{PML}}$ statistic. Whether or not this is a good strategy, and which values of q and τ should in practice be used to implement it, are open questions which would deserve a further investigation.

To conclude, note that our proposed testing strategy could be used for evaluating a much broader class of models that the special class of models considered here. It may indeed actually be used for checking the distributional aspect of virtually any parametric conditional models with continuous (univariate) dependent variable.

To see this, simply observe that for any correctly specified and dynamically complete[29] model for the conditional densities $f_t(y_t|x_t;\theta)$ of some dependent variable y_t given some information set x_t, according to the lemma given in Diebold et al. (1998) to which we referred to in Section 3, the conditional probability integral transform $v_t=v_t(y_t,x_t,\theta)=F_t(y_t|x_t;\theta)$ of the observations of the model, where $F_t(y_t|x_t;\theta)$ denotes the conditional cdf. associated with the conditional pdf. $f_t(y_t|x_t;\theta)$, must be independent of x_t and identically and independently distributed as a continuous uniform random variable over [0,1] Accordingly, the distributional aspect of any such parametric conditional models may thus be checked, as we did it for our generic class of models[30], through a m-test verifying if the sample moments of the estimated conditional probability integral transform $\hat{v}_t=v_t(y_t,x_t,\hat{\theta}_n)=F_t(y_t|x_t;\hat{\theta}_n)$ of the observations of the model are in accordance with those of the continuous uniform distribution, and that using the same m-test statistic[31] $\mathcal{M}_n^{\mathrm{PML}}$ or $\mathcal{M}_n^{\mathrm{OPG}}$. Further, dynamic distributional misspecification could likewise be checked by looking at autocorrelations in $\hat{v}_t$ as suggested above.

Whether or not the automatic application of this general procedure provides a sensible and useful diagnostic test depends on the model at hand. It clearly provides a useful diagnostic test in all models where, as in the class of models considered here, the distributional aspect of the model is well separated from the other aspects of the specification of the model. Examples of such models in the financial area are the generalized GARCH-type model considered by Hansen (1994), where the shape parameters of the postulated innovations density are allowed to vary as a parametric function of the information set x_t, and the class of autoregressive conditional duration models recently introduced by Engle and Russel (1998). We believe that it may also provide a useful generally applicable diagnostic test in other models. For example, again in the financial area, it could be used for evaluating the specification of stochastic volatility models after estimation by simulated maximum likelihood, the conditional probability integral transform of the observations of the model being itself computed by simulation techniques. The exploration of the empirical relevance and usefulness of these potential extensions is on our agenda.

References

Andrews, D.W.K., 1988a. Chi-square diagnostic tests for econometric models: introduction and applications. Journal of Econometrics 37, 135–156.

Andrews, D.W.K., 1988b. Chi-square diagnostic tests for econometric models: theory. Econometrica 56 (6), 1419–1453.

Bai, Jushan, 2003. Testing parametric conditional distributions of dynamic models. The Review of Economics and Statistics 85, 531–549.

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., 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., Nelson, B., 1994. In: Engle, R., McFadden, D. (Eds.), Arch models. Ch. 49 in Handbook of Econometrics, vol. 4. North Holland, Amsterdam.

Engle, R.F., Russel, J.R., 1998. Autoregressive conditional duration: a new model for irregularly spaced transaction data. Econometrica 66 (5), 1127–1163.

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

Diebold, F.X., Gunther, T.A., Tay, A.S., 1998. Evaluating density forecasts with applications to financial risk management. International Economic Review 39, 863–883.

Ding, Z., Granger, C.W.J., Engel, R.F., 1993. A long memory property of stock market returns and a new model. Journal of Empirical Finance 1, 83–106.

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

Giot, P., Laurent, S., 2003. Value-at-risk for long and short trading positions. Journal of Applied Econometrics 18, 641–663.

Gourieroux, C., Monfort, A., 1995. Statistics and Econometric Models. Cambridge University Press, Cambridge.

Gourieroux, C., Monfort, A., Trognon, A., 1984. Pseudo-maximum likelihood methods: theory. Econometrica 52 (3), 681–700.

Hansen, B.E., 1994. Autoregressive conditional density estimation. International Economic Review 35, 705–730.

[28] Note that, contrary to standard ML, the Gaussian PML estimator of the model is still consistent in this more general situation.

[29] On this concept, see Wooldridge (1994).

[30] In our generic class of models, the conditional probability integral transform of the observations of the model simply corresponds to the (unconditional) probability integral transform of the innovations of the model.

[31] Note that the Gauss procedure to which we referred in footnote 12 may also be used for this more general purpose.

Inoue, Atsushi (1998), "A conditional goodness-of-fit test for time series", Mimeo, Department of Agricultural and Resource Economics, North California State University, Raleigh, USA.
Jarque, C.M., Bera, A.K., 1980. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics Letters 6, 255–259.
Johnson, N.L., Kotz, S., 1970. Distributions in Statistics: Continuous Univariate Distributions - 2. John Wiley and Sons, New-York.
Kim, S., Shephard, N., Chib, S., 1998. Stochastic volatility: likelihood inference and comparison with ARCH models. Review of Economic Studies 65, 361–393.
Lambert, Ph., Laurent, S., 2001. Modelling financial time series using GARCH-type models with a skewed Student distribution for the innovations. Discussion Paper no 01–25, Institut de Statistique, Université Catholique de Louvain, Louvain-la-Neuve, Belgium.
Lye, J.N., Martin, V.L., Teo, L.E., 1998. Parametric distributional flexibility and conditional variance models with application to hourly exchange rates. Working Paper no 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 specification testing and conditional moment tests. Econometrica 53 (5), 1047–1070.
Palm, F.C., 1996. In: Maddala, G.S., Rao, C.R., Vinod, H.D. (Eds.), GARCH models of volatility. Ch. 17 in Handbook of Statistics, vol. 14. North Holland, Amsterdam.
Paolella, M.C., 1999. Using flexible GARCH models with asymmetric distributions. Working paper, Institute of Statistics and Econometrics, Christian Albrechts University, Kiel, Germany.
Rosenblatt, M., 1952. Remarks on multivariate transformation. Annals of Mathematical Statistics 23, 470–472.
Tauchen, G., 1985. Diagnostic testing and evaluation of maximum likelihood models. Journal of Econometrics 30, 415–443.
Whang, Y-J., Andrews, D.W.K., 1993. Tests of specification for parametric and semiparametric models. Journal of Econometrics 57, 277–318.
White, H., 1987. In: Bewley, T. (Ed.), Specification testing in dynamic models. Ch. 1 in Advances in Econometrics — Fifth World Congress, vol. 1. Cambridge University Press, New-York.
White, H., 1994. Estimation, Inference and Specification Analysis. Cambridge University Press, Cambridge.
Wooldridge, J., 1990. A unified approach to robust, regression-based specification 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. Specification testing and quasi-maximum likelihood estimation. Journal of Econometrics 48, 29–55.
Wooldridge, J., 1994. In: Engle, R., McFadden, D. (Eds.), Estimation and inference for dependent processes. Ch. 45 in Handbook of Econometrics, vol. 4. North Holland, Amsterdam.