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

Bernard Lejeune

University of Liège, ERUDITE and CORE

Plan

1- Introduction

2- Model specification and estimation

3- Testing distributional specification through moments of probability integral transform

4- Test statistics

5- Monte-Carlo evidence

1. Introduction

- Background:
  - Parametric conditional heteroscedasticity models, as the standard Student t GARCH model, are a customary tool for analyzing financial data.
  - If these models are routinely estimated in applications, diagnostic testing of their specification, and more particularly of their distributional specification, is in practice much less common.

- Objective:
  - To provide a *convenient* and *generally applicable* diagnostic test for checking the distributional aspect of these models.

- First idea:
  - By analogy to the popular Jarque-Bera (1980) test for normality, checking through a m-test that the third and fourth order sample moments of the (estimated) innovations of the model are in accordance with their (estimated) theoretical values.

  → *Convenient* since m-test are standard and easy to implement, **but** not *generally applicable* because it requires existence of moments up to order eight (unlikely when working with a number of popular models such as the standard Student t GARCH model).

- To overcome this problem while staying in the convenient m-testing framework, this paper suggests:
  - A m-test based, instead of the moments of the innovations themselves, on the moments of the *probability integral transform* (i.e. cdf. transform) of the innovations.

  → Characteristics:
  - (relatively) easy to implement.
  - generally applicable.
  - well-behaved both in terms of size and power.

2. Model specification and estimation

- Notation :
  - y_t : (continuous) dependent variable of interest.
  - z_t : vector of explanatory variables.
  - x_t : information set $x_t \equiv (z_t, y_{t-1}, z_{t-1}, ..., y_1, z_1)$. If no explanatory variables, $x_t \equiv (y_{t-1}, ..., y_1)$.

- Model specification :

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

where - γ is a vector of parameters,
- $\mu_t(.,.)$ and $h_t(.,.) > 0$ are known scalar functions,
- ε_t are i.i.d. zero mean and unit variance innovations independent of x_t with density $g(\varepsilon; \eta)$, where η is a vector of shape parameters.

- This specification defines a fully parametric model $\mathcal{P}$ for the conditional densities of y_t given x_t :

$$\mathcal{P} \equiv \left\{ f_t(y_t|x_t; \theta) = \frac{1}{\sqrt{h_t(x_t, \gamma)}}\, g\left(\frac{y_t - \mu_t(x_t, \gamma)}{\sqrt{h_t(x_t, \gamma)}}; \eta \right) : \theta = (\gamma', \eta')' \in \Theta \right\}$$

whose, by construction, $E(y_t|x_t) = \mu_t(x_t, \gamma)$ and $V(y_t|x_t) = h_t(x_t, \gamma)$, $t = 1, 2, ...$

- In typical applications :
  - $\mu_t(x_t, \gamma)$ is specified according to an AR, MA or ARMA process.
  - $h_t(x_t, \gamma)$ is specified according to some autoregressive scheme such as ARCH, GARCH, EGARCH, ...
  - $g(\varepsilon; \eta)$ is chosen among standardized continuous distributions allowing for fat tails and possibly further for asymmetry.

- A customary example: the pure time-series Student $t(\nu)$ AR(1)-GARCH(1,1) model obtained by setting

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

and

$$g(\varepsilon;\eta) = b g^*(b\varepsilon;\nu)$$

where $u_{t-1} = y_{t-1} - \mu_{t-1}(x_{t-1},\gamma)$, $h_{t-1} = h_{t-1}(x_{t-1},\gamma)$, $b = \sqrt{\frac{\nu}{\nu-2}}$ and $g^*(w;\nu)$ is the usual Student t density with ν degrees of freedom.

- Maximum likelihood estimator:

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

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, if model $\mathcal{P}$ is correctly specified, i.e. if there exists some true value θ^o in Θ such that

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

where $p_t^o(y_t|x_t)$ denotes the true conditional density of y_t given x_t, the ML estimator $\hat{\theta}_n$ yields a consistent, efficient and asymptotically normal estimator of the unknown true value $\theta^o = (\gamma^{o\prime}, \eta^{o\prime})'$ of $\mathcal{P}$:

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

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

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

3. Testing distributional specification through moments of probability integral transform

- We consider testing:

  $$\mathrm{H}_0: \text{ model } \mathcal{P} \text{ is correctly specified}$$

  against

  $$\mathrm{H}_1: \text{ model } \mathcal{P} \text{ is misspecified due to distributional misspecification}$$

  It is implicitly assumed that the conditional mean and variance specifications have successfully been checked in a previous stage.

- By analogy to Jarque-Bera (1980), a natural strategy would be to check through a m-test that the misspecification indicator

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

  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.

- Statistical rationale of this strategy:
  - Under H_0, $\hat{\theta}_n \to \theta^o$ and $E\left[(e_t^o)^r - \phi_r(\eta^o)\right] = 0$, $t = 1, 2, ...$
  - Under H_1, $\hat{\theta}_n \to \theta_n^*$ and (usually) $E\left[(e_t^*)^r - \phi_r(\eta_n^*)\right] \neq 0$, $t = 1, 2, ...$

- Problem of this strategy:
  - To be applicable, it requires that $E\left[\left((e_t^o)^4 - \phi_r(\eta^o)\right)^2\right] < \infty$, i.e. that under H_0, at the true value $\theta^o = (\gamma^{o\prime}, \eta^{o\prime})'$, the assumed density $g(\varepsilon; \eta^o)$ possesses finite moments up to order 8.

  $\rightarrow$ This is beyond what we can expect to be fulfilled in applications when working with a number popular models (e.g. the standard Student t GARCH model).

- Proposed alternative strategy :
  - To check the moments of a (judiciously chosen) transformation of the (estimated) innovations rather than the moments of the (estimated) innovations themselves.

- Probability integral transform of the estimated innovations $\hat{e}_t$:

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

where

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

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

- Regardless of $g(\varepsilon; \eta)$, under H_0, $v_t^o = v_t(y_t, x_t, \theta^o) = G(e_t^o; \eta^o)$ must be independent of x_t and identically and independently distributed as a continuous uniform r.v. over $[0, 1]$, whose central moments (all finite) are

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

- It follows that :
  - Under H_0, we must have $E\left[(v_t^o - 0.5)^r - \delta(r)\right] = 0$, $t = 1, 2, ...$
  - Under H_1, we will (usually) have $E\left[(v_t^* - 0.5)^r - \delta(r)\right] \neq 0$, $t = 1, 2, ...$

- This suggests checking through a m-test the closeness to zero of a $(q \times 1)$ misspecification indicator of the form

$$\hat{M}_n = \frac{1}{n} \sum_{t=1}^{n} \begin{bmatrix} \hat{v}_t - 0.5 \\ (\hat{v}_t - 0.5)^2 - \delta(2) \\ \vdots \\ (\hat{v}_t - 0.5)^q - \delta(q) \end{bmatrix}$$

→ This strategy is applicable without any restriction on the existence of the moments of the true innovations and for any choice of q.

→ Theoritically, setting $q = 2$ already allows to detect departures from the assumed density both in terms of skewness and kurtosis.

4. Test statistics

- Given the assumed statistical setup, under general regularity conditions, a proper m-test statistic for checking the closeness to zero of the $q \times 1$ misspecification indicator

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

is given by the asymptotically chi-square statistic

$$\mathcal{M}_n = n\hat{M}_n \hat{K}_n^{-1} \hat{M}_n \xrightarrow{d} \chi^2(q)$$

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

$$\begin{aligned} K_n^o &= \frac{1}{n}\sum_{t=1}^{n} E\left[\left(m_t^o - D_n^o A_n^{o-1} s_t^o\right)\left(m_t^o - D_n^o A_n^{o-1} s_t^o\right)'\right] \\ &= \frac{1}{n}\sum_{t=1}^{n} E\left[m_t^o m_t^{o\prime}\right] - \frac{1}{n}\sum_{t=1}^{n} E\left[m_t^o s_t^{o\prime}\right]\left(\frac{1}{n}\sum_{t=1}^{n} E\left[s_t^o s_t^{o\prime}\right]\right)^{-1} \frac{1}{n}\sum_{t=1}^{n} E\left[s_t^o m_t^{o\prime}\right] \end{aligned}$$

where

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

- Remarks :
  - The equality of the two expressions of K_n^o follows from the so-called information matrix (i.e. $A_n^o = -\frac{1}{n}\sum_{t=1}^{n} E\left[s_t^o s_t^{o\prime}\right]$) and cross-information matrix ($D_n^o = -\frac{1}{n}\sum_{t=1}^{n} E\left[m_t^o s_t^{o\prime}\right]$) equalities.
  - Numerous consistent estimators of K_n^o are conceivable, but only few (essentially the two ones outlined hereafter) have the highly desirable property to always deliver at least semi-positive definite (and usually positive definite) estimates.

- The simplest operational form of $\mathcal{M}_n$ is obtained by taking as a consistent estimator of K_n^o

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

This yields

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

which in practice may be computed as n minus the residual sum of squares ($= nR_u^2$) of the OLS artificial regression

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

→ The statistic $\mathcal{M}_n^{OPG}$ is particularly easy to implement. Unfortunately, it is well-known for often exhibiting (very) poor finite sample properties (tendency to over-reject when the null is true).

- An interesting alternative statistic is obtained by taking as a consistent estimator of K_n^o

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

This yields

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

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

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

→ If somewhat less computationally convenient, the statistic $\mathcal{M}_n^{PML}$ is usually (much) better behaved in finite sample than the statistic $\mathcal{M}_n^{OPG}$.

5. Monte-Carlo Evidence

- Questions:
  - Is the proposed distributional m-testing strategy effective?
  - What is the best way to implement it in practice?

→ Simulation study of the finite sample performance of six versions of the proposed test (its $\mathcal{M}_n^{OPG}$ and $\mathcal{M}_n^{PML}$ forms with $q = 2$, 4 and 6) for checking the distributional specification of two models.

- Considered models:

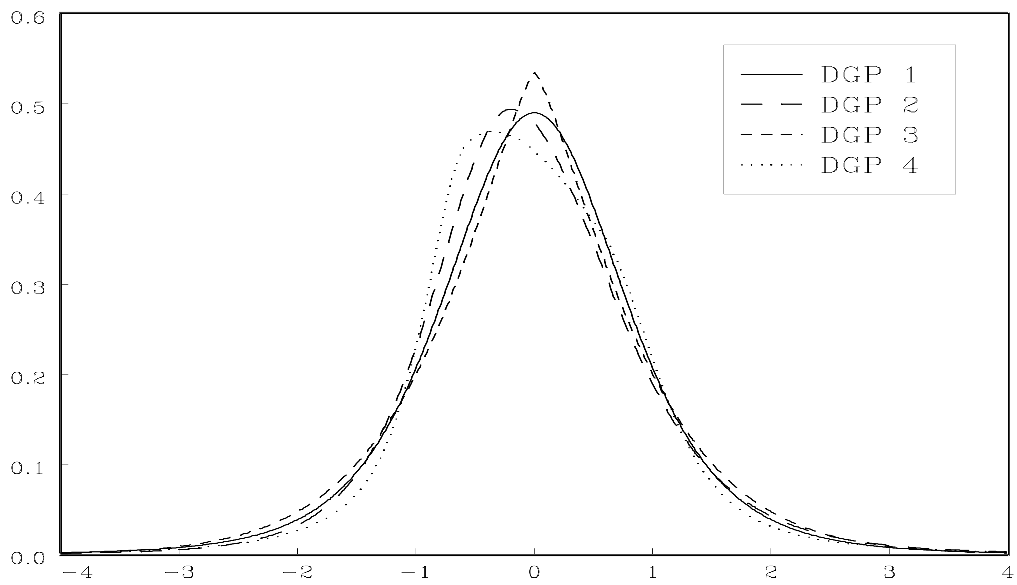
$$y_t = \gamma_1 + \gamma_2 y_{t-1} + \sqrt{\gamma_3 + \gamma_4 u_{t-1}^2 + \gamma_5 h_{t-1}}\, \varepsilon_t \qquad (1)$$

with - Model 1: $\varepsilon_t \sim$ (standardized) Student $t(\nu)$
- Model 2: $\varepsilon_t \sim$ (standardized) skewed Student $t(\nu, \kappa)$

- Considered DGP:

Equation (1) where $\gamma_1 = 0, \gamma_2 = 0.1, \gamma_3 = 0.05, \gamma_4 = 0.1, \gamma_5 = 0.8$

with - 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) GED(1.3)
- DGP 4: $\varepsilon_t \sim$ mixture of two (standardized) skewed Student t



- Monte-Carlo results (tests at 5%, 5000 / 2000 replications) :

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)

- Tests size :
  - OPG tests are systematically over-sized (size range : $[5.7\%, 17.2\%]$).
  - PML tests are in all cases pretty well-sized (size range : $[3.0\%, 6.1\%]$).

$\longrightarrow$ Unless n is large and q small, the $\mathcal{M}_n^{PML}$ statistic should be preferred.

- Tests power :
  - Size-corrected power of OPG and PML tests are similar.
  - Setting $q=2$ is not enough and $q=6$ does not seem to pay off.

$\longrightarrow$ Setting $q=4$ seems to be ‘the best’ and appears to ensure ‘good’ power against various alternatives.