OPTIMAL PREDICTIONS OF POWERS OF CONDITIONALLY HETEROSKEDASTIC PROCESSES

C. Francq, J.-M. Zakoïan
University Lille 3,
Villeneuve d'Ascq, France
e-mail: christian.francq@univ-lille3.fr

Abstract

The standard method for estimating powers of conditionally heteroskedastic processes is a two-step procedure in which the volatility is estimated by gaussian quasi-maximum likelihood (QML) in a first step, and an empirical mean of the rescaled innovations is computed in a second step. This paper proposes an alternative one-step procedure, based on an appropriate non-gaussian QML estimation of the model, and establishes the asymptotic properties of the two approaches. Their performances are compared for finite-order GARCH models and for the ARCH(∞). For the standard GARCH(p, q) and the Asymmetric Power GARCH(p, q), it is shown that the asymptotic relative efficiency of the estimators only depends on the prediction problem and on some moments of the independent process. An application to indexes of major stock exchanges is proposed.

1 Introduction

Most of the conditional heteroscedastic models proposed in the literature, in particular finite-order standard or asymmetric GARCH as well as ARCH(∞), can be embedded in the model

$$\begin{cases} \epsilon_t = \sigma_t \eta_t \\ \sigma_t = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots; \theta_0) \end{cases} \tag{1}$$

where (η_t) is an iid sequence of random variables, with η_t independent of $\{\epsilon_u, u < t\}$, $\theta_0 \in \mathbb{R}^m$ is a parameter belonging to a parameter space Θ, and $\sigma : \mathbb{R}^\infty \times \Theta \to (0, \infty)$. For evident identifiability reasons, a scale constraint is required on the sequence (η_t). The standard assumption is $E\eta_t^2 = 1$ but any other constraint of the form $E|\eta_t|^r = 1$, with $r \neq 0$, can be used as well (provided that the r-th moment exists).

For any real number r such that $E|\eta_t|^r < \infty$, the optimal predictor, in the L^2 sense, of $|\epsilon_t|^r$ given its entire past is

$$E_{t-1}(|\epsilon_t|^r) = \sigma_t^r E|\eta_t|^r,$$

where E_{t-1} denotes expectation conditional on the infinite past. Given observations $(\epsilon_1, \ldots, \epsilon_n)$, we consider two approaches for predicting $|\epsilon_{n+1}|^r$ with $r \neq 0$:

- A *mixed* (parametric and non parametric) two-step approach in which θ_0 is estimated under the usual assumption that $E|\eta_t|^2 = 1$ and $E|\eta_t|^r$ is estimated non-parametrically. The prediction of $|\epsilon_{n+1}|^r$ is then the estimated value of σ_{n+1}^r multiplied by the estimate of $E|\eta_t|^r$.

- A *fully parametric* one-step approach in which θ_0 is estimated under the assumption that $E|\eta_t|^r = 1$. The prediction of $|\epsilon_{n+1}|^r$ is then simply the estimated value of σ_{n+1}^r.

The mixed approach is standard. The fully parametric approach is new, to our knowledge. We also adapt the two approaches for predicting $\log|\epsilon_t|$. With the mixed two-step approach, the optimal prediction is $E_{t-1}(\log|\epsilon_t|) = \log\sigma_t + E\log|\eta_t|$, provided that $E\log|\eta_t|$ exists. With the fully parametric one-step approach, the optimal prediction is $E_{t-1}(\log|\epsilon_t|) = \log\sigma_t$, under the identifiability constraint $E\log|\eta_t| = 0$. This prediction problem will be referred to as at the case $r = 0$ because, via the Box-Cox transformation $\log|x| = \lim_{r\to 0}(|x|^r - 1)/r$, it can be considered as the limit of the prediction of $|\epsilon_t|^r$ when r tends to zero. We establish the asymptotic properties of the two approaches and compare their efficiencies.

2 One-step prediction

Given $\theta \in \Theta$ and arbitrary initial values $\tilde{\epsilon}_i$ for $i \leq 0$, a proxy of the volatility is defined by $\tilde{\sigma}_t(\theta) = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_1, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \ldots; \theta)$. Given an *instrumental* density h, generalized QMLE are defined in [1] by

$$\hat{\theta}_{n,h} = \arg\max_{\theta\in\Theta} \frac{1}{n}\sum_{t=1}^{n} g(\epsilon_t, \tilde{\sigma}_t(\theta)), \qquad g(x,\sigma) = \log\frac{1}{\sigma}h\left(\frac{x}{\sigma}\right).$$

This estimator is the standard gaussian QMLE when h is the standard gaussian density ϕ. Under the identifiability constraint $E|\eta_t|^r = 1$ when $r \neq 0$, and $E\log|\eta_t| = 0$ when $r = 0$, we show that the generalized QMLE is consistent if the instrumental density h is chosen in the class $\mathcal{C}(r)$ of functions of the form

$$\begin{cases} c|x|^{\lambda-1}\exp\left(-\lambda|x|^r/r\right), & \text{if } r > 0, \\ c|x|^{-\lambda-1}\exp\left(\lambda|x|^r/r\right), & \text{if } r < 0, \\ \sqrt{\lambda/\pi}|2x|^{-1}\exp\left\{-\lambda(\log|x|)^2\right\}, & \text{if } r = 0, \end{cases}$$

for constants $\lambda, c > 0$. More precisely, it is shown that, under mild regularity conditions and when $h \in \mathcal{C}(r)$,

$$\sqrt{n}\left(\hat{\theta}_{n,h} - \theta_0\right) \xrightarrow{\mathcal{L}} \mathcal{N}(0, 4\tau_{h,f}^2 J^{-1})$$

where

$$J = J(\theta_0) = E\left(\frac{1}{\sigma_t^4}\frac{\partial\sigma_t^2}{\partial\theta}\frac{\partial\sigma_t^2}{\partial\theta'}(\theta_0)\right) \quad \text{and} \quad \tau_{h,f}^2 = \frac{Eg_1^2(\eta_0,1)}{\left\{Eg_2(\eta_0,1)\right\}^2},$$

with $g_1(x,\sigma) = \partial g(x,\sigma)/\partial\sigma$ and $g_2(x,\sigma) = \partial g_1(x,\sigma)/\partial\sigma$. With the direct approach based on the generalized QMLE $\hat{\theta}_{n,h}$, the optimal prediction $E_n\left(|\epsilon_{n+1}|^r\right)$ is thus estimated by

$$P_{n,h} = \tilde{\sigma}^r\left(\epsilon_n, \epsilon_{n-1}, \ldots; \hat{\theta}_{n,h}\right).$$

3 Two-step prediction

Assume that there exists a function F such that for any $\theta \in \Theta$, for any $K > 0$, and any $(x_i)_i$

$$K\sigma(x_1, x_2, \ldots; \theta) = \sigma(x_1, x_2, \ldots; \theta^*), \quad \text{where } \theta^* = F(\theta, K).$$

Standard GARCH with volatility $\sigma_t^2 = \omega + \sum_{i=1}^{q} \alpha_i \epsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2$ obviously verify this assumption with

$$F(\theta, K) = (K^2\omega, K^2\alpha_1, \ldots, K^2\alpha_q, \beta_1, \ldots, \beta_p)'.$$

Let $\theta_0^* = F(\theta_0, \sqrt{\mu_2})$ where $\mu_s = E|\eta_t|^s$ for $s \neq 0$. Model (1) with $E|\eta_t|^r = 1$ can be reparameterized as

$$\begin{cases} \epsilon_t = \sigma_t^* \eta_t^*, \quad E\eta_t^{*2} = 1, \\ \sigma_t^* = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots; \theta_0^*) \end{cases}$$

The gaussian QMLE of θ_0^* is defined by

$$\hat{\theta}_n^* = \arg\max_{\theta \in \Theta} \frac{1}{n} \sum_{t=1}^{n} \log \frac{1}{\tilde{\sigma}_t(\theta)} \phi\left(\frac{\epsilon_t}{\tilde{\sigma}_t(\theta)}\right).$$

Let the rescaled residuals

$$\hat{\eta}_t^* = \frac{\epsilon_t}{\hat{\sigma}_t^*}, \quad \text{where } \hat{\sigma}_t^* = \sigma(\epsilon_{t-1}, \epsilon_{t-2}, \ldots, \tilde{\epsilon}_0, \tilde{\epsilon}_{-1}, \ldots; \hat{\theta}_n^*)$$

and, for $r \neq 0$, let

$$\hat{\mu}_r^* = \frac{1}{n} \sum_{t=1}^{n} |\hat{\eta}_t^*|^r, \quad \mu_r^* = E|\eta_t^*|^r = \frac{1}{\mu_2^{r/2}}, \quad \kappa_s = \frac{E|\eta_t|^s}{\mu_2^{s/2}}.$$

Under some regularity conditions, the joint asymptotic distribution of the QMLE and $\hat{\mu}_r^*$ is given by

$$\begin{pmatrix} \sqrt{n}\left(\hat{\theta}_n^* - \theta_0^*\right) \\ \sqrt{n}(\hat{\mu}_r^* - \mu_r^*) \end{pmatrix} \xrightarrow{\mathcal{L}} \mathcal{N}\left\{0, \begin{pmatrix} (\kappa_4 - 1)J_*^{-1} & -\lambda_r J_*^{-1}\Omega_* \\ -\lambda_r \Omega_*' J_*^{-1} & \sigma_{\mu_r^*}^2 \end{pmatrix}\right\},$$

where

$$J_* = E\left(\frac{1}{\sigma_t^{*4}} \frac{\partial \sigma_t^2(\theta_0^*)}{\partial \theta} \frac{\partial \sigma_t^2(\theta_0^*)}{\partial \theta'}\right), \quad \Omega_*' = E\left(\frac{1}{\sigma_t^{*2}} \frac{\partial \sigma_t^2(\theta_0^*)}{\partial \theta'}\right)$$

and $\lambda_r = \frac{r}{2}\kappa_r(\kappa_4 - 1) - (\kappa_{2+r} - \kappa_r)$, $\sigma_{\mu_r^*}^2 = \kappa_{2r} - \kappa_r^2 + \frac{r}{2}\kappa_r(\lambda_r - \kappa_{2+r} + \kappa_r)$. It follows that, for $r \neq 0$, the two-step estimation of the optimal predictor $E_n\left(|\epsilon_{n+1}|^r\right)$ is given by

$$P_n^* = \tilde{\sigma}^r(\epsilon_n, \epsilon_{n-1}, \ldots; \hat{\theta}_n^*)\hat{\mu}_r^* = \tilde{\sigma}^r(\epsilon_n, \epsilon_{n-1}, \ldots; \tilde{\theta}_n),$$

where the asymptotic distribution of $\tilde{\theta}_n = F(\hat{\theta}_n^{*'}, \{\hat{\mu}_r^*\}^{1/r})$ is easily obtained from the previous results, by the delta method. In the case $r = 0$, *i.e.* for predicting $\log|\epsilon_{n+1}|$, the asymptotic behavior of the two-step method is obtained similarly.

4 Comparison

It can be shown that for the GARCH, or for the larger class of the APARCH models, the one-step method is asymptotically more efficient than the two-step one, in the sense that the difference between the asymptotic variances of $\tilde{\theta}_n$ and $\hat{\theta}_{n,h}$ is semi-definite positive, if and only if

$$\left(\frac{2}{r}\right)^2\left(\frac{\kappa_{2r}}{\kappa_r^2}-1\right) \geq \kappa_4 - 1$$

when $r \neq 0$ and

$$4\mathrm{Var}(\log|\eta_0|) \geq \kappa_4 - 1$$

when $r = 0$. Surprisingly, this efficiency criterion only involves r and moments of the iid noise, but not θ_0, which allows to select straightforwardly the more efficient method, in function of r and estimated moments of η_t.

Figure 1 shows the ARE of the one-step method relative to the two step method as measured by the ratios

$$(\kappa_4 - 1)/\left(\frac{2}{r}\right)^2\left(\frac{\kappa_{2r}}{\kappa_r^2}-1\right) \quad \text{when } r \neq 0, \text{ and } \frac{\kappa_4 - 1}{4\mathrm{Var}(\log|\eta_0|)} \quad \text{when } r = 0$$

for Student distributed η_t. It is seen that the one-step method outperforms the indirect one when $r \in (0.5, 2)$. On the contrary, for $r > 2$ and small or negative values of r, the two-step approach is preferable. The differences are particularly spectacular for small value of ν. The ARE's are displayed as dots in the case $r = 0$. The ARE are continuous functions of r because, one can show that, in a quite general framework,

$$\left(\frac{2}{r}\right)^2\left(\frac{\kappa_{2r}}{\kappa_r^2}-1\right) \to 4\mathrm{Var}(\log|\eta_0|) \quad \text{as } r \to 0.$$

We also show that, in particular for the prediction of the absolute value of daily stock market indices, the one-step method is typically more efficient (and simpler) that the naive two-step method. The preprint of the paper is available in [2].

References

[1] Berkes I., Horváth L. (2004). The efficiency of the estimators of the parameters in GARCH processes. *The Annals of Statistics*. Vol. **32**, pp. 633-655.

[2] Francq C., Zakoïan J.-M. (2010). Optimal predictions of powers of conditionally heteroskedastic processes. *Munich Personal RePEc Archive (MPRA)*. Paper No. **22155**.

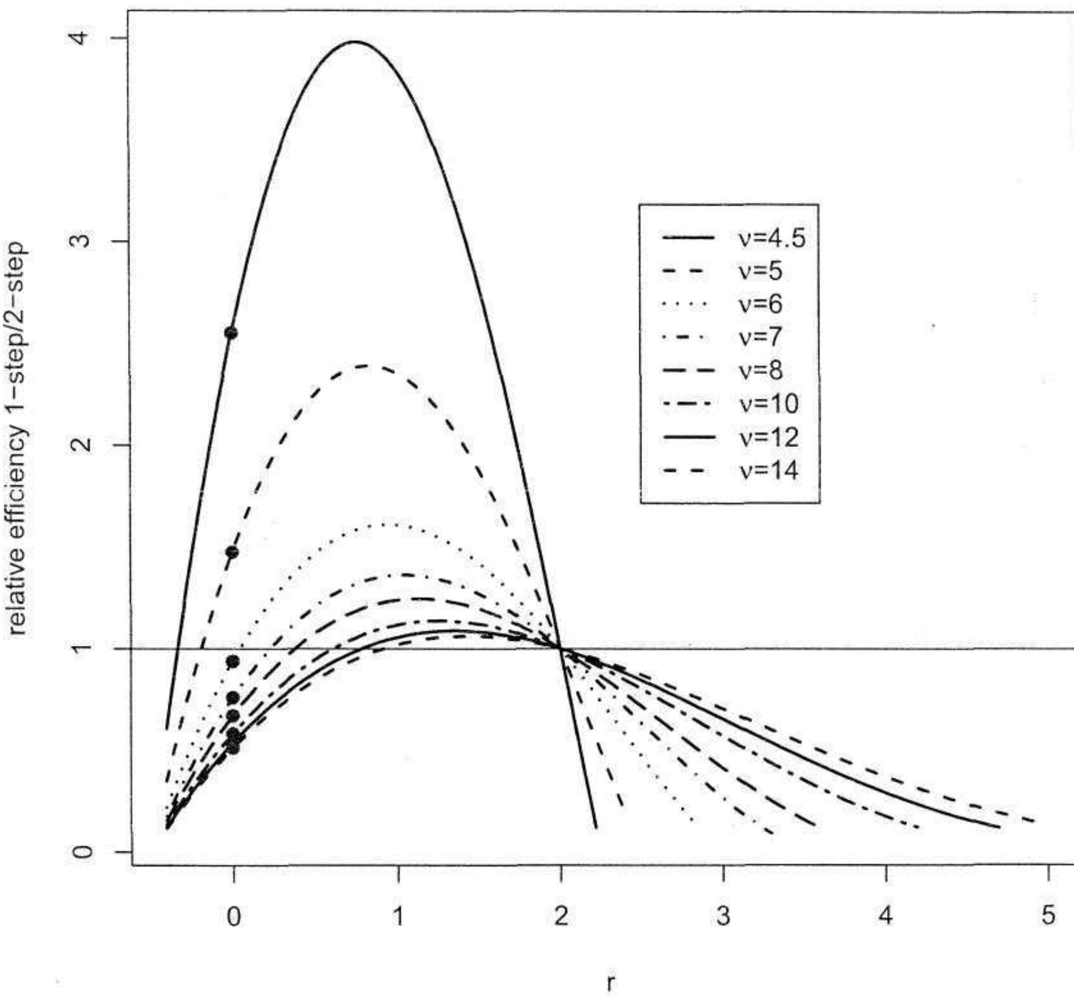

Figure 1: Asymptotic relative efficiency of the one-step method relative to the two step method for Student distributions with parameter ν.