Estimation of Parametric Models with Conditional Heteroscedastic Errors

Christian Lavergne, Yann Vernaz

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

Estimation Of Parametric Models With Conditional Heteroscedastic Errors

Christian Lavergne, Yann Vernaz

N° 3658

Avril 1999

THÈME 4

Rapport de recherche

ISSN 0249-6399

Estimation Of Parametric Models With Conditional Heteroscedastic Errors

Christian Lavergne, Yann Vernaz

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet IS2

Rapport de recherche n° 3658 — Avril 1999 — 28 pages

Abstract: We consider a model with conditional heteroscedastic errors. The model requires only the form of conditional mean and conditional variance functions to be specified. We propose an effective approach for fitting this class of model. Our estimator is deduced from quasi-likelihood concept using an iterative and adaptive procedure. The convergence properties are established. Finally, our method and widely used estimators are compared via numerical experiments.

Key-words: Conditional Heteroscedasticity; Likelihood; Pseudo-likelihood; Quasi-likelihood; Least Squares; Adaptive Method.

(Résumé : tsvp)

Unité de recherche INRIA Rhône-Alpes
655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN (France)
Téléphone : 04 76 61 52 00 - International: +33 4 76 61 52 00
Télécopie : 04 76 61 52 52 - International: +33 4 76 61 52 52

Estimation dans les modèles avec des erreurs conditionnellement hétéroscédastiques

Résumé : Nous considérons des modèles avec des erreurs conditionnellement hétéroscédastiques. Nous proposons une procédure d'estimation efficace en deux étapes. La première étape, basée sur le concept de quasi-vraisemblance, permet d'obtenir un estimateur consistant dans un contexte très général. L'efficacité asymptotique est atteinte par une procédure adaptative. Notre procédure et les principales méthodes existantes sont comparées par des simulations de Monte-Carlo.

Mots-clé : conditionnellement hétéroscédastique, vraisemblance, pseudo-vraisemblance, quasi-vraisemblance, moindres carrés, estimation adaptative.

Introduction

In the analysis of certain time series, the class of ARMA process can be proved unsuited. We can quote monetary and financial phenomenon analysis, which have some specificity which can not be take accounted by a traditional ARMA modeling. Indeed, they are characterized by non-linear dynamic and marked volatility (or instantaneous variability). To design the volatility, Engle [5] proposed an autoregressive representation of conditional variance function to its last information. This model is called ARCH (Auto-Regressive Conditional Heteroscedastic). The general principle suggested by Engle allows the variance to depend on the available information by a specification where the square of disturbances follows an autoregressive process. This idea have been generalized. The purpose of this paper is to propose a simple and efficient algorithm to evaluate the parameters of a model with conditional heteroscedastic errors.
From a statistical point of view, these models constitute a specific class of non-linear model that can be completely studied. Much work has been completed on the subject, and for a review see Bollerslev, Cabbage and Kroner [3] and Bera and Higgins [1].
This article is structured as follows. Section 1 outlines considered models pointing out main definitions and assumptions. In Section 2 we examine methods based on maximum likelihood principle, their properties and limitations are discussed. In Section 3 we study least squares procedures in a semi-parametric framework. In Section 4 we propose and study a method based on quasi-likelihood concept. This concept authorizes a statistical inference by appointing only the first two moments and the function which connects them. The asymptotic efficiency is obtained by building an adaptive procedure. Finally, all estimators are compared via numerical experiments in Section 5.

1 The Model

1.1 Definitions

Let $\varepsilon_t(\theta)$ denote a discrete time stochastic process defined on a probabilized space $(\Omega, \mathcal{F}, P)$ with conditional mean and variance functions parameterized with a finite dimensional vector $\theta \in \Theta \subseteq \mathbf{R}^m$, θ_0 denoting the true value. For notational simplicity we shall assume that $\varepsilon_t(\theta)$ is a scalar. Also, let $E(./\mathcal{F}_{t-1})$ denote the mathematical expectation, conditional on the past of the process, $\mathcal{F}_{t-1}$ denoting the information set available at $t-1$.

Definition 1.1 *The $\varepsilon_t(\theta_0)$ process follows a Conditional Heteroscedastic (CH) model if the conditional mean equals zero but the conditional variance,*

$$\sigma_t^2(\theta_0) = E(\varepsilon_t^2(\theta_0)/\mathcal{F}_{t-1}), \quad t = 1, 2, ...,$$

depends on the sigma-field generated by the past observations.

For example, the original Engle's [5] ARCH(p) model satisfies the following random difference equation:

$$\varepsilon_t(\theta_0) = u_t \sigma_t(\theta_0) \quad \text{with} \quad \sigma_t^2(\theta_0) = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2. \tag{1}$$

We have $\theta = (\alpha_0, \alpha_1, ..., \alpha_p)$ with $\alpha_0 > 0$, $\alpha_i \geq 0$ for $i = 1, \cdots, p$, $(u_t)_{t \in \mathbf{Z}}$ being iid white noise independent of $\{\varepsilon_s, s < t\}$ with $E(u_t) = 0$ and $E(u_t^2) = 1$. It is easily to prove that $E(\varepsilon_t(\theta_0)/\mathcal{F}_{t-1}) = 0$ and $E(\varepsilon_t^2(\theta_0)/\mathcal{F}_{t-1}) = \sigma_t^2(\theta_0)$.

Condition for existence of $2m$ moments is translated for each ARCH model by a condition on parameters α_i, $i = 1, ..., p$. For example, an ARCH(1) model admits $2m$ moments if and only if $\alpha_1^m \prod_{j=1}^{m} (2j-1) < 1$. Then, if a finite second moment exists the variance is equal to:

$$E(\varepsilon_t^2) = \frac{\alpha_0}{1 - \sum_{i=1}^{p} \alpha_i}.$$

Thus, an ARCH model is homoscedastic but conditionally heteroscedastic. In addition, for any k and $h > 0$ we have $E(\varepsilon_t \varepsilon_{t-k}) = 0$ and $E(\varepsilon_t \varepsilon_{t-k}/\mathcal{F}_{t-h}) = 0$. Finally, an ARCH($p$) process is strictly stationary if $\sum_{i=1}^{p} \alpha_i < 1$ (*cf.* Engle [5]).

Now we extend the study to the situation in which $\varepsilon_t(\theta_0)$ corresponds to the innovations from some more elaborate models. This allows to introduce various additional effects of explanatory variables, either in conditional mean or variance functions.

Definition 1.2 *The stochastic process y_t define a model with CH errors if the conditional mean equal $m_t(\theta_0)$ and conditional variance equal $\sigma_t^2(\theta_0)$ >0, for $t = 1, 2, ...$*

Let us define the process $\varepsilon_t(\theta_0) = y_t - m_t(\theta_0)$ for $t = 1, 2, ...$, then the conditional variance of $\varepsilon_t(\theta_0)$ is equal to the conditional variance of the process y_t. Moreover, the standardized process $u_t(\theta_0) = \varepsilon_t(\theta_0)\sigma_t^{-1}(\theta_0)$ will have conditional mean zero, and a time invariant conditional variance of unity. To avoid any ambiguity we shall restrict our attention to parameterization

$$y_t = m_t(\theta_0) + \sigma_t(\theta_0) u_t. \tag{2}$$

Notice that the linear mean function can be an autoregressive moving average model or any type of regression model.

1.2 Assumptions

$\theta \in \Theta$, where Θ is compact and has a nonempty interior.

Assumption 1. $m_t(\theta)$ (linear) and $\sigma_t(\theta)$ are known twice continuously differentiable functions with respect to the components of $\theta \in \mathrm{Int}\Theta$.
The process σ_t^2 is strictly stationary, ergodic and bounded below by $\sigma > 0$. Moreover, we suppose that $E(\sigma_t^4) < +\infty$.

Assumption 2. The two matrix

$$M(\theta) = E\left(\frac{\partial m_t(\theta)}{\partial\theta}\frac{\partial m_t(\theta)}{\partial\theta'}\right), \quad \Sigma(\theta) = E\left(\frac{\partial \sigma_t^2(\theta)}{\partial\theta}\frac{\partial \sigma_t^2(\theta)}{\partial\theta'}\right)$$

converge to a positive definite matrix.

Assumption 3. The random variables u_t are i.i.d. independent of $\sigma_t(\theta)$ with symmetric, absolutely continuous Lebesgue density f, where $f(x) > 0 \ \forall x \in \mathbf{R}$, such as

$$I_j = \int \Psi_j^2(x) f(x) dx < +\infty, \quad \text{for } j = 1, 2 \tag{3}$$

with

$$\Psi_1(x) = \frac{f'(x)}{f(x)} \quad \text{and} \quad \Psi_2(x) = 1 + x\frac{f'(x)}{f(x)}.$$

The density f is generally assumed to be Gaussian, but in this work f is not necessarily known. Such processes, in the particular case where the conditional variance is modeled by an ARCH process, was introduced by Engle and Gonzalez-Rivera [6].

2 Maximum Likelihood Inference

In this section we examine estimators built on the likelihood principle. The construction of a likelihood function is based on an assumed conditional probability density function (PDF). We can estimate the parameters with a maximum likelihood (ML) method where the conditional PDF is known. If the conditional PDF is unknown we use a pseudo-maximum likelihood (PML) method where conditional normality is assumed, even though this assumption may be false.

2.1 The conditional density is known

Let us consider the model (2). For a sample of length T, the log-likelihood function is given by

$$\mathcal{L}_T(\theta) = \sum_{t=1}^{T} \log\left[\frac{f(u_t(\theta))}{\sigma_t(\theta)}\right]. \tag{4}$$

If the probability density function f is known, the log-likelihood function is entirely determined and the ML estimator (MLE) is found by maximizing equation (4) with respect of parameter θ. Then, the MLE, $\hat{\theta}_{ml}$, solves the system of equations

$$\sum_{t=1}^{T}\left[\frac{1}{\sigma_t(\theta)}\frac{\partial m_t(\theta)}{\partial\theta}\frac{f'(u_t(\theta))}{f(u_t(\theta))}+\frac{1}{2\sigma_t^2(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\left(1+u_t(\theta)\frac{f'(u_t(\theta))}{f(u_t(\theta))}\right)\right] = 0 \qquad (5)$$

where f' is the derivative of $f(.)$ respect to u_t and $u_t(\theta)=(y_t-m_t(\theta))/\sigma_t(\theta)$.

Consistency and asymptotic normality properties of MLE are only established for a limited class of CH process, mainly ARCH(p), see Weiss [22], Lumsdaine [15] and Lee and Hansen [14]. Under correct specification of the conditional variance function, the ergodic theorem and a central limit theorem can be invoked to state that

$$\sqrt{T}\left(\hat{\theta}_{ml}-\theta_0\right)\longrightarrow\mathcal{N}(0,J(\theta_0)^{-1}) \text{ with } J(\theta_0)=-E_0\left(\frac{\partial^2\mathcal{L}_T(\theta_0)}{\partial\theta\partial\theta'}\right).$$

E_0 indicates mathematical expectation taken compared to the true distribution. The existence and strict positivity of matrix $J(\theta_0)$ results in assumptions 1,2 and 3.

2.2 The conditional density is unknown

In such a case, we cannot apply a ML procedure. The idea is to seek a ML method applied as if the u_t process were a Gaussian white noise. It is shown whereas this error of possible specification on the law does not have a consequence on the convergence properties of the PML estimator. For a detailed theoretical study of the PML method useful references are White [23], Gourieroux, Monfort and Trognon [10] and Gourieroux and Monfort [9].

By definition, the PMLE of θ is found by maximizing equation (4) under the assumption of conditional normality. This assumption results in $f'(u_t)/f(u_t)=-u_t$ and $1+u_tf'(u_t)/f(u_t)=1-u_t^2$, consequently equation (5) becomes:

$$\sum_{t=1}^{T}\left[\frac{\partial m_t(\theta)}{\partial\theta}\frac{u_t(\theta)}{\sigma_t(\theta)}+\frac{1}{2\sigma_t^2(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\left(u_t^2(\theta)-1\right)\right] = 0. \qquad (6)$$

Under regularity conditions discussed in Wooldridge [24], Lee and Hansen [14], Lumsdaine [15] and Weiss [22] the limiting distribution of the PMLE is,

$$\sqrt{T}\left(\hat{\theta}_{pml}-\theta_0\right)\longrightarrow\mathcal{N}\left(0,J(\theta_0)^{-1}I(\theta_0)J(\theta_0)^{-1}\right) \text{ with } I(\theta_0)=E_0\left(\frac{\partial\mathcal{L}_T(\theta_0)}{\partial\theta}\frac{\partial\mathcal{L}_T(\theta_0)}{\partial\theta'}\right).$$

Remarks: The PMLE is less efficient than the MLE, reflecting the lack of information about the PDF. However, if the true distribution is normal, I and J coincide, the PMLE is asymptotically efficient. The finite sample properties of PMLE end efficiency loss in regard to the MLE have been studied in several Monte Carlo simulations by Weiss [22], Bollerslev and Wooldridge [4] and Engle and Gonzàlez-Rivera [6]. The loss of efficiency can be proven very significant. A study of Gonzàlez-Rivera [8] showed that the loss of efficiency between the PMLE and the MLE is proportional to the kurtosis coefficient and Fisher's information matrix.

Engle and Gonzàlez-Rivera [6] introduced a third methodology denoted semi-parametric (SP) estimation. In this situation, the PDF is assumed sufficiently smooth to be approximated by a non-parametric procedure. The SP estimator is a two-step estimator. In the first step, consistent estimates of the parameters of interest are obtained through PML and used to construct a non-parametric density of the standardized innovations. The second step consists of using this non-parametric density to design a non-parametric likelihood that is maximized with respect to θ. The SP procedure is more effective than the PML method, but remains less effective than ML method. Moreover, the SP procedure can require a significant computing time.

Generally, equations (5) and (6) are highly non-linear in θ, the solution is achieved by numerical techniques, involving numerical difficulties. In particular, the likelihood can be maximized using the algorithm described in the article of Berndt, Hall, Hall and Hausman [2], commonly called BHHH. He has proved to be suitable for this class of model.

3 Least Squares Method

As shown by Weiss [22] and Pantula [18], least squares estimators for ARCH models are consistent and asymptotically normal, but less efficient than ML estimator to estimate the conditional mean. Engle [5] in his founder article uses an ordinary least squares (OLS) method to estimate the parameters of a regression model with ARCH errors. Pantula [18] proposed a quasi-generalized least squares (QGLS) method for an AR model with ARCH noise. He showed that the QGLS estimator can be less effective, but that the asymptotic loss of efficiency is bounded below. Let us note that all the authors consider only a limited class of CH process in a Gaussian framework. In this section we study the least squares methods in a more general setting. The proof of various results are put in Appendix A.

3.1 Two Steps Estimator

We use the representation $\varepsilon_t^2 = \sigma_t^2(\theta) + \eta_t$ for the process (1) where $E(\eta_t/\mathcal{F}_{t-1}) = 0$ and $E(\eta_t^2/\mathcal{F}_{t-1}) = \sigma_t^4 E(u_t^2 - 1)^2$. If u_t is a Gaussian white noise then $E(\eta_t^2/\mathcal{F}_{t-1}) = 2\sigma_t^4$.

The least squares method allows two principal stages, which correspond to the two levels of the model. We pose $\theta = (b, \alpha)$ where b denoting the conditional mean parameter and α

the conditional variance parameter. A first step gives the OLS estimators by maximizing $\sum_{t=1}^{T} \varepsilon_t^2$ and $\sum_{t=1}^{T} \eta_t^2$ with $\eta_t = \tilde{\varepsilon}_t^2 - \sigma_t^2(\theta)$ and $\tilde{\varepsilon}_t = y_t - m_t(\hat{b}_{ols})$.

Proposition 3.1 *$\hat{b}_{ols}$ and $\hat{\alpha}_{ols}$ converge almost surely to b_0 and α_0 and*

$$\sqrt{T}\left(\hat{b}_{ols} - b_0\right) \longrightarrow \mathcal{N}(0, P(\theta_0)), \quad \sqrt{T}\left(\hat{\alpha}_{ols} - \alpha_0\right) \longrightarrow \mathcal{N}(0, B(\theta_0)),$$

where

$$P(\theta_0) = E_0\left(\frac{\partial m_t(\theta_0)}{\partial \theta}\frac{\partial m_t(\theta_0)}{\partial \theta'}\right)^{-1} E_0\left(\sigma_t^2(\theta_0)\frac{\partial m_t(\theta_0)}{\partial \theta}\frac{\partial m_t(\theta_0)}{\partial \theta'}\right) E_0\left(\frac{\partial m_t(\theta_0)}{\partial \theta}\frac{\partial m_t(\theta_0)}{\partial \theta'}\right)^{-1}$$

and

$$B(\theta_0) = E_0(u_t^2 - 1)^2 A(\theta_0)^{-1} E_0\left(\sigma_t^4(\theta_0)\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'}\right) A(\theta_0)^{-1}$$

with

$$A(\theta_0) = E_0\left(\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'}\right).$$

We have $E_0(u_t^2-1)^2 = E_0(u_t^4) - 1 = k - 1$ where $k = E_0(u_t^4)/E_0(u_t^2)^2 = E_0(u_t^4)$ denoting the kurtosis.

The OLS method does not take account of the heteroscedasticity phenomenon. We can deduce a conditional variance estimate by compute $\tilde{\sigma}_t^2 = \sigma_t^2(\hat{\theta}_{ols})$. The new estimator is calculate by using the QGLS method by maximizing $\displaystyle\sum_{t=1}^{T} \frac{\varepsilon_t^2}{\tilde{\sigma}_t^2}$ and $\displaystyle\sum_{t=1}^{T} \frac{\eta_t^2}{\tilde{\sigma}_t^4}$ with $\eta_t = \tilde{\varepsilon}_t^2 - \sigma_t^2(\theta)$ and $\tilde{\varepsilon}_t = y_t - m_t(\hat{b}_{qgls})$.

Proposition 3.2 *$\hat{b}_{qgls}$ and $\hat{\alpha}_{qgls}$ converge almost surely to b_0 and α_0 and*

$$\sqrt{T}\left(\hat{b}_{qgls} - b_0\right) \longrightarrow \mathcal{N}(0, Q^{-1}(\theta_0)), \quad \sqrt{T}\left(\hat{\alpha}_{qgls} - \alpha_0\right) \longrightarrow \mathcal{N}(0, C(\theta_0)^{-1}),$$

where

$$Q(\theta_0) = E_0\left(\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial \theta}\frac{\partial m_t(\theta_0)}{\partial \theta'}\right)$$

and

$$C(\theta_0) = \frac{1}{k-1} E_0\left(\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'}\right).$$

The MLE asymptotic covariance matrix is $J(\theta_0)^{-1}$ with

$$J(\theta_0) = I_1 E_0\left(\frac{1}{2\sigma_t^4(\theta_0)}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'} + \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial \theta}\frac{\partial m_t(\theta_0)}{\partial \theta'}\right).$$

If the PDF is normal $J(\theta_0)$ becomes

$$J_N(\theta_0) = E_0\left(\frac{1}{2\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'} + \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial\theta}\frac{\partial m_t(\theta_0)}{\partial\theta'}\right).$$

The PMLE asymptotic covariance matrix is $J_N(\theta_0)^{-1}I(\theta_0)J_N(\theta_0)^{-1}$ with

$$\begin{aligned} I(\theta_0) &= E_0\left(\frac{1}{4\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}(k-1) + \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial\theta}\frac{\partial m_t(\theta_0)}{\partial\theta'}\right) \\ &+ E_0\left(\frac{1}{2\sigma_t^3(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}u_t^3(\theta_0)\right). \end{aligned}$$

Corollary 3.1 *For a Gaussian CH process (i.e. $\varepsilon_t = \sigma_t(\theta_0)u_t$ with $u_t \sim \mathcal{N}(0,1)$), QGLS, MLE and PMLE are asymptotically equivalent. If u_t is non-Gaussian then QGLS and PMLE are asymptotically equivalent and the efficiency losses with respect to the MLE are proportional to $I_1(k-1)/4$.*

Proof: The proof comes immediately from comparison between above asymptotic matrix in the case $m_t(\theta) = 0$. We have

$$C(\theta_0) = \frac{1}{k-1}E_0\left(\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right),$$

$$J(\theta_0) = I_1 E_0\left(\frac{1}{2\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right)$$

and

$$J_N(\theta_0)I(\theta_0)^{-1}J_N(\theta_0) = \frac{1}{k-1}E_0\left(\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'}\right).$$

So, we deduce the following immediate results: If $u_t \sim \mathcal{N}(0,1)$, then $I_1 = 1$ and $k-1 = 2$ from where $C(\theta_0) = J(\theta_0) = J_N(\theta_0)I(\theta_0)^{-1}J_N(\theta_0)$. Otherwise $C(\theta_0) = J_N(\theta_0)I(\theta_0)^{-1}J_N(\theta_0)$ and $J(\theta_0) = \dfrac{(k-1)I_1}{2}J_N(\theta_0)I(\theta_0)^{-1}J_N(\theta_0)$. We conclude by using $(k-1)I_1 \geq 2$.

Corollary 3.2 *If $u_t \sim \mathcal{N}(0,1)$, the QGLS asymptotic matrix of the conditional mean parameter is more important than PMLE.*

Proof: The proof is immediate by comparison between

$$Q(\theta_0) = E_0\left(\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial\theta}\frac{\partial m_t(\theta_0)}{\partial\theta'}\right)$$

and

$$J_N(\theta_0)I(\theta_0)^{-1}J_N(\theta_0) = E_0\left(\frac{1}{2\sigma_t^4(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta'} + \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial\theta}\frac{\partial m_t(\theta_0)}{\partial\theta'}\right).$$

3.2 Application to an ARCH process

Using $\eta_t = \varepsilon_t^2 - \sigma_t^2(\theta)$, equation (1) is re-writed in the following form:

$$\varepsilon_t^2 = \alpha_0 + \sum_{i=1}^{p} \alpha_i \varepsilon_{t-i}^2 + \eta_t.$$

Thus we obtain an AR(p) representation on the square of ε_t. We pose $\Phi_t = (1, \varepsilon_{t-1}^2, \cdots, \varepsilon_{t-p}^2)$ and $\theta = (\alpha_0, ..., \alpha_p)$. Then we introduce naturally a least squares method while regressing ε_t^2 on $1, \varepsilon_{t-1}^2, \cdots, \varepsilon_{t-p}^2$. The OLS estimator is calculated explicitly by:

$$\hat{\theta}_{ols} = \left(\sum_{t=1}^{T} \Phi_t' \Phi_t \right)^{-1} \sum_{t=1}^{T} \Phi_t' \varepsilon_t^2.$$

According to proposition 3.1, $\sqrt{T}\left(\hat{\theta}_{ols} - \theta_0\right) \longrightarrow \mathcal{N}(0, B(\theta_0))$

$$\text{where } B(\theta_0) = (k-1) E_0(\Phi_t'\Phi_t)^{-1} E_0(\sigma_t^4(\theta_0)\Phi_t'\Phi_t) E_0(\Phi_t'\Phi_t)^{-1}.$$

The second step corresponds to the following estimator

$$\hat{\theta}_{qgls} = \left(\sum_{t=1}^{T} \frac{\Phi_t'\Phi_t}{\tilde{\sigma}_t^4} \right)^{-1} \sum_{t=1}^{T} \frac{\Phi_t'\varepsilon_t^2}{\tilde{\sigma}_t^4},$$

and according to proposition 3.2,

$$\sqrt{T}\left(\hat{\theta}_{qgls} - \theta_0\right) \longrightarrow \mathcal{N}(0, C(\theta_0)^{-1}) \text{ with } C(\theta_0) = \frac{1}{k-1} E_0\left(\frac{\Phi_t'\Phi_t}{\sigma_t^4(\theta_0)}\right).$$

3.3 Regression Model with ARCH errors

For the regression model, $y_t = x_t b + \sigma_t(\theta) u_t$, the first step is given by:

$$\hat{b}_{ols} = \left(\sum_{t=1}^{T} x_t' x_t \right)^{-1} \sum_{t=1}^{T} x_t' y_t \quad \text{and} \quad \hat{\alpha}_{ols} = \left(\sum_{t=1}^{T} \tilde{\Phi}_t' \tilde{\Phi}_t \right)^{-1} \sum_{t=1}^{T} \tilde{\Phi}_t' \tilde{\varepsilon}_t^2$$

with

$$\tilde{\varepsilon}_t = y_t - x_t \hat{b}_{ols} \text{ and } \tilde{\Phi}_t = (1, \tilde{\varepsilon}_{t-1}^2, \cdots, \tilde{\varepsilon}_{t-p}^2).$$

We have:

$$\sqrt{T}\begin{pmatrix} \hat{b}_{ols} - b \\ \hat{\alpha}_{ols} - \alpha \end{pmatrix} \longrightarrow \mathcal{N}\left(0, \begin{bmatrix} P(\theta_0) & 0 \\ 0 & B(\theta_0) \end{bmatrix}\right)$$

where

$$P(\theta_0) = E_0(x_t'x_t)^{-1} E_0(\sigma_t^2(\theta_0) x_t' x_t) E_0(x_t'x_t)^{-1}.$$

This first step allows to deduce a conditional variance estimator: $\tilde{\sigma}_t^2 = \tilde{\Phi}_t \hat{\alpha}_{ols}$. The second step used the QGLS method, *i.e.*

$$\hat{b}_{qgls} = \left(\sum_{t=1}^{T} \frac{x_t' x_t}{\tilde{\sigma}_t^2} \right)^{-1} \sum_{t=1}^{T} \frac{x_t' y_t}{\tilde{\sigma}_t^2} \text{ and } \hat{\alpha}_{qgls} = \left(\sum_{t=1}^{T} \frac{\tilde{\Phi}_t' \tilde{\Phi}_t}{\tilde{\sigma}_t^4} \right)^{-1} \sum_{t=1}^{T} \frac{\tilde{\Phi}_t' \tilde{\varepsilon}_t^2}{\tilde{\sigma}_t^4}.$$

We have:

$$\sqrt{T} \begin{pmatrix} \hat{b}_{qgls} - b \\ \hat{\alpha}_{qgls} - \alpha \end{pmatrix} \longrightarrow \mathcal{N} \left(0, \begin{bmatrix} Q(\theta_0) & 0 \\ 0 & C(\theta_0) \end{bmatrix}^{-1} \right)$$

where

$$Q(\theta_0) = E_0 \left(\frac{x_t' x_t}{\sigma_t^2(\theta_0)} \right).$$

The QGLS estimator of conditional mean parameter is asymptotically less effective than the PMLE (see Corollary 3.2). It is important to specify this loss of effectiveness. We have studied by simulation the relative asymptotic effectiveness from QGLS ratio to PML for a regression model with ARCH errors when $\sum_{i=1}^{p} \alpha_i$ is close to 1. This numerical study shows that relative asymptotic effectiveness is $\geq 1/3$. Pantula [18] establish a similar result for an AR(1) process with ARCH(1) errors. He shows that the relative asymptotic effectiveness is bounded below by $1/(1+2\alpha_1)$.

In short: The QGLS and PMLE of the conditional variance part are asymptotically equivalent. However, the QGLS estimator does not have an optimal property for the conditional mean parameter but the asymptotic loss of effectiveness is bounded below.

4 Quasi-Likelihood Method

To estimate the parameters the statistics use mainly two methods of inference, the least squares and the ML methods. Godambe and Heyde's [7] work unified those two approaches under a general description called quasi-likelihood (QL). In this section we show that, the QL method is well adapted for models with CH errors.

4.1 A Brief Introduction

Let $y_t, t = 1, 2, ..., T$ denote a discrete stochastic process. We focus on models which the error distribution has only its first and second moment properties specified. The objective is the optimal estimation of the parameter of interest θ. Accordingly, we define the class of the estimating functions with null mean and square integrable such as:

$$\mathcal{G} = \{ G_T(\{y_t, t = 1, ..., T\}, \theta) \text{ and } E(G_T(\theta) G_T(\theta)') \text{ no singular} \}.$$

By definition, we call quasi-likelihood estimator (QLE) the solution of the equation:

$$G_T(\hat{\theta}_{ql}) = 0.$$

Subsequently, we always place in a subspace $\mathcal{H} \subseteq \mathcal{G}$. If $G_T(\theta) \in \mathcal{H}$ we define the standardized estimating functions

$$G_T^{(s)}(\theta) = -\left(E\frac{\partial G_T(\theta)}{\partial \theta}\right)' (EG_T(\theta)G_T(\theta)')^{-1}G_T(\theta).$$

Then, the variance is equal to:

$$\mathrm{Var}\left[G_T^{(s)}(\theta)\right] = \left(E\frac{\partial G_T(\theta)}{\partial \theta}\right)' E\left(G_T(\theta)G_T(\theta)'\right)^{-1}\left(E\frac{\partial G_T(\theta)}{\partial \theta}\right).$$

The concept of optimality within $\mathcal{H}$ is carried out by maximizing the covariance matrix of $G_T^{(s)}(\theta)$. We can introduce an information criterion,

$$\mathcal{E}(G_T(\theta)) = E(G_T^{(s)}(\theta)G_T^{(s)}(\theta)')$$

who is a generalization of Fisher's information. We say that $G_T^\star$ is an optimal estimating function (or a quasi-score function) in the class $\mathcal{H}$ if

$$\mathrm{Var}\left[G_T^{\star(s)}(\theta)\right] \geq \mathrm{Var}\left[G_T^{(s)}(\theta)\right], \quad \forall G_T(\theta) \in \mathcal{H}.$$

A fundamental theorem, established by Heyde [12], allows to characterize a quasi-score function.

Theorem 4.1 *$G_T^\star(\theta) \in \mathcal{H}$ is a quasi-score function if*

$$E\left(G_T^{\star(s)}(\theta)G_T^{(s)'}(\theta)\right) = E\left(G_T^{(s)}(\theta)G_T^{\star(s)'}(\theta)\right) = E\left(G_T^{(s)}(\theta)G_T^{(s)'}(\theta)\right) \tag{7}$$

otherwise

$$\left(E\frac{\partial G_T(\theta)}{\partial \theta}\right)^{-1} E\left(G_T(\theta)G_T^\star(\theta)'\right) \tag{8}$$

is a constant matrix for all $G_T \in \mathcal{H}$. Conversely, if $\mathcal{H}$ is convex and $G_T^\star(\theta) \in \mathcal{H}$ is a quasi-score function, then relations (7) and (8) are verified.

Our motivation is the optimal estimation of the parameters in a model with CH errors. Now, we propose to build a quasi-score function for this class of models.

4.2 Model with CH Errors and Quasi-Likelihood

Let us consider the model (2). Then, for the class of estimating functions

$$\mathcal{H} = \left\{H \;:\; H_T(\theta) = \sum_{t=1}^{T} a_t(\theta)\left(y_t - m_t(\theta)\right), \;\; a_t \text{ is } \mathcal{F}_{t-i} \text{ measurable}\right\},$$

we have

$$\mathrm{Var}(H_T^{(s)}(\theta)) = \sum_{t=1}^{T} \left(a_t(\theta) \frac{\partial m_t(\theta)}{\partial \theta} \right) \left(\sum_{t=1}^{T} a_t(\theta) a_t(\theta)' \sigma_t^2(\theta) \right)^{-1} \sum_{t=1}^{T} \left(a_t(\theta) \frac{\partial m_t(\theta)}{\partial \theta} \right)',$$

who is maximized by using the Cauchy-Schwarz's inequality if:

$$a_t(\theta) = k(\theta) \frac{\partial m_t(\theta)}{\partial \theta} \sigma_t^{-2}(\theta), \quad t = 1, 2, ..., T,$$

where $k(\theta)$ is an unspecified multiplier. Finally, the quasi-score function is written:

$$H_T^{\star}(\theta) = \sum_{t=1}^{T} \frac{\partial m_t(\theta)}{\partial \theta} \sigma_t^{-2}(\theta)(y_t - m_t(\theta)). \tag{9}$$

So, the QLE is solution of the system $H_T^{\star}(\hat{\theta}_{ql}) = 0$.

4.3 Wedderburn's Quasi-Likelihood

Consider a conditional variance function σ_t^2, which does not depend on θ. We consider the space of estimating functions $\mathcal{H} = \{a_t(y_t - m_t(\theta))\}$, where a_t does not depend on θ. The following result due to Wedderburn [21] allows to define a quasi-score function within $\mathcal{H}$.

Proposition 4.1 *The estimate function*

$$G_T^{\star}(\theta) = \sum_{t=1}^{T} \frac{\partial m_t(\theta)}{\partial \theta} \sigma_t^{-2}(\theta)(y_t - m_t(\theta))$$

is a quasi-score function within $\mathcal{H}$.

The proof is immediate starting from the theorem 4.1.

This quasi-score function is widely used in practice. In particular, to estimate the parameters of a generalized linear model (*cf.* McCullagh and Nelder [17]). This quasi-score function satisfies the usual properties of the log-likelihood function:

$$E(G_T(\theta)) = 0 \text{ and } E\left(\frac{\partial G_T(\theta)}{\partial \theta}\right) = -E(G_T(\theta) G_T(\theta)').$$

The solution of $G_T(\hat{\theta}_{ql}) = 0$ is obtained easily by numerical techniques. The equations to solve correspond to the QGLS equations where the variance σ_t^2 is re-actualized with each new value of the parameters by using QGLS method. In fact, we apply an iterative re-weighted QGLS procedure.

The asymptotic properties of this estimator are due to McCullagh [16]. Under adequate assumptions, we have:

$$\sqrt{T}(\hat{\theta}_{ql} - \theta_0) \xrightarrow{law} \mathcal{N}\left(0, \Gamma(\theta_0)^{-1}\right) \quad \text{with } \Gamma(\theta_0) = E_0\left(-\frac{\partial G_T(\theta)}{\partial \theta}\right).$$

Proposition 4.2 *If $u_t \sim \mathcal{N}(0,1)$, the QLE is asymptotically equivalent to the PMLE.*

Proof: For a model with CH errors, the asymptotic covariance matrix of PMLE is

$$\left[E_0\left(\frac{1}{2\sigma_t^4(\theta_0)}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'} + \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial \theta}\frac{\partial m_t(\theta_0)}{\partial \theta'}\right)\right]^{-1}.$$

By using $E(u_t^2) = 1$ and $\partial \sigma_t^2(\theta)/\partial\theta = 2\sigma_t(\theta)\partial\sigma_t(\theta)/\partial\theta$ we show that

$$\begin{aligned}
E_0\left(-\frac{\partial G_T(\theta_0)}{\partial \theta}\right) &= E_0\left(-\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}\frac{\partial m_t(\theta_0)}{\partial \theta}\varepsilon_t + \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial \theta}\frac{\partial m_t(\theta_0)}{\partial \theta'}\right) \\
&= E_0\left(-\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}\frac{\partial (y_t - \sigma_t(\theta_0)u_t)}{\partial \theta}\sigma_t(\theta_0)u_t + \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial \theta}\frac{\partial m_t(\theta_0)}{\partial \theta'}\right) \\
&= E_0\left(\frac{1}{\sigma_t^4(\theta_0)}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}\frac{\partial \sigma_t(\theta_0)}{\partial \theta}\sigma_t(\theta_0) + \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial \theta}\frac{\partial m_t(\theta_0)}{\partial \theta'}\right) \\
&= E_0\left(\frac{1}{2\sigma_t^4(\theta_0)}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta}\frac{\partial \sigma_t^2(\theta_0)}{\partial \theta'} + \frac{1}{\sigma_t^2(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial \theta}\frac{\partial m_t(\theta_0)}{\partial \theta'}\right).
\end{aligned}$$

4.4 Adaptive Procedure

The goal is to recapture the asymptotic efficiency losses of the QLE to MLE. For this, we propose an adaptive procedure. An estimator such as $\hat{\theta}_T = \theta_0 + o_p(1)$ means that the difference $\hat{\theta}_T - \theta_0$ is an infinitely small in probability, *i.e.* $\forall \varepsilon > 0,\ \ P(\|\hat{\theta}_T - \theta_0\| > \varepsilon) = 0$. $\hat{\theta}_T$ is termed effective if he is asymptotically equivalent to the MLE, *i.e.*

$$\sqrt{T}(\hat{\theta}_T - \theta_0) = I(\theta_0, f)^{-1}\Delta_T(\theta_0, f) + o_p(1)$$

where

$$\Delta_T(\theta, f) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\partial l_t(\theta)}{\partial \theta} \quad \text{and} \quad l_t(\theta) = \log f(u_t(\theta)) - \frac{1}{2}\log \sigma_t^2(\theta).$$

We suppose the PDF known. Under traditional regularity conditions the MLE of θ_0 is consistent, asymptotically normal with a covariance matrix $I(\theta_0, f)^{-1}$. Let $\hat{\theta}_T$ denote a $\sqrt{T}$-consistent estimator of θ_0, then we construct the estimator

$$\hat{\hat{\theta}}_T = \hat{\theta}_T + \frac{1}{\sqrt{T}}\hat{I}_T(\hat{\theta}_T, f)^{-1}\Delta_T(\hat{\theta}_T, f),$$

where the matrix $I(\theta_0, f)$ is estimated by:

$$\hat{I}_T(\hat{\theta}_T, f) = \frac{1}{T}\sum_{t=1}^{T} \frac{\partial l_t(\hat{\theta}_T)}{\partial \theta} \frac{\partial l_t(\hat{\theta}_T)}{\partial \theta'}$$

and

$$\Delta_T(\hat{\theta}_T, f) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T} \frac{\partial l_t(\hat{\theta}_T)}{\partial \theta}.$$

If $\hat{\theta}_T$ is a $\sqrt{T}$-consistent estimator of θ_0, then $\hat{I}_T(\hat{\theta}_T, f)$ is a consistent estimator of $I(\theta_0, f)$. A proof of this result can be found in Bollerslev and Wooldridge [4].

$$\hat{I}(\hat{\theta}_T, f) = I(\theta_0, f) + o_p(1) \tag{10}$$

Under some regularity conditions, see Section 1, and if θ_T is a $\sqrt{T}$-consistent estimator of θ_0 then

$$\Delta_T(\hat{\theta}_T, f) = \Delta_T(\theta_0, f) - I(\theta_0, f)\sqrt{T}(\hat{\theta}_T - \theta_0) + o_p(1) \tag{11}$$

The proof of this result put in Appendix B.

Then, from (10) and (11) we establish

Proposition 4.3 *$\hat{\hat{\theta}}_T$ is an effective estimator of θ_0.*

Indeed, we have

$$\begin{aligned}
\sqrt{T}(\hat{\hat{\theta}}_T - \theta_0) &= \sqrt{T}(\hat{\theta}_T - \theta_0) + \hat{I}_T(\hat{\theta}_T, f)^{-1}\Delta_T(\hat{\theta}_T, f) \\
&= \sqrt{T}(\hat{\theta}_T - \theta_0) + I(\theta_0, f)^{-1}\Delta_T(\hat{\theta}_T, f) + o_p(1) \\
&= \sqrt{T}(\hat{\theta}_T - \theta_0) + I(\theta_0, f)^{-1}\left[\Delta_T(\theta_0, f) - I(\theta_0, f)\sqrt{T}(\hat{\theta}_T - \theta_0)\right] + o_p(1) \\
&= I(\theta_0, f)^{-1}\Delta_T(\theta_0, f) + o_p(1)
\end{aligned}$$

Thus, if the PDF is known we obtain an effective estimator starting from a consistent estimator of θ_0 and by carrying out an iteration of the Newton-Raphson algorithm.

Finally, we propose the following procedure:

- *Step 1*: Choose some initial consistent estimates of the set of parameters θ_0. These estimates may come from applying QGLS.
- *Step 2*: Use the QL method to estimate θ.
- *Step 3*: If the density is known, perform the QL method with adaptive procedure.

5 Monte Carlo Simulations

To test how the procedures performs in terms of efficiency, we have carried out Monte Carlo simulations for three types of models, a process ARCH(1), $\varepsilon_t = u_t\sigma_t$ where $\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2$, a process ARCH(3), $\varepsilon_t = u_t\sigma_t$ where $\sigma_t^2 = \alpha_0 + \alpha_1\varepsilon_{t-1}^2 + \alpha_2\varepsilon_{t-2}^2 + \alpha_3\varepsilon_{t-3}^2$ and a regression model with ARCH errors. For each process, we have considered that the disturbance term u_t is distributed as:

- a Gaussian distribution

$$f(u_t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u_t^2}{2}\right),$$

- or as a gamma distribution with shape parameter s,

$$f(u_t) = \frac{\sqrt{s}}{\Gamma(s)} \left(\sqrt{s}u_t + s\right)^{s-1} \exp\left(\sqrt{s}u_t + s\right)$$

with $E(u_t) = 0$ and $E(u_t^2) = 1$. The coefficient of skewness is $2/\sqrt{s}$ and the coefficient of kurtosis $3 + 6/s$. When s tends to infinity, it converges to the normal distribution (Johnson and Kotz [13]).

All the estimations results are based on 100 replications. For the ARCH parameter we took various values to study the evolution of the estimations. Note that the unconditional fourth moment does not exist in any of the models we considers[1].

The sample sizes varies between 300 and 500. For each model, we calculated the average and the standard deviation of the estimates. The convergence of the estimators is appreciated for an accuracy of 10^{-4} with a maximum of 200 iterations. Beyond that, we consider that there is not convergence. In the same way, we announce the cases where the estimates are apart from the parameters space (i.e. $\alpha_0 \leq 0$ and $\sum_{i=1}^{p} \alpha_i > 1$). We note these situations by "No". The studied methods are OLS, QGLS, QL, Adaptive QL, ML and PML.

5.1 Study on Gaussian simulations

Tables 1 and 2: As awaited the OLS method, which does not take into account of heteroscedasticity phenomenon, is the least effective method. The use of the QGLS procedure makes a improvement, but QL and ML methods give the best numerical results. The results obtained by these two methods are identical, indeed the equations to be solved are the same ones. Let us note that algorithm BHHH becomes very sensitive to the initial values for extreme values of the parameters (*i.e.* when the sum of the parameters is close to 1). We can notice that the QGLS and QL procedures give results systematically when the algorithm of BHHH converges. The reciprocal one being false. Finally, let us announce that the QL procedure

[1] For example an ARCH(1) model with $\alpha_1 > 0.5774$ have not fourth moment.

is much faster in calculating times.

Table 3: The estimate of the parameters of conditional variance, $\alpha = (\alpha_0, ..., \alpha_p)$, is numerically the same for the three methods; QL, Adaptive QL and ML. On the other hand, the QGLS procedure is less effective, particularly for the extreme values of the parameters. With regard to the estimate of the regression parameter, methods AQL and ML are most powerful. Nevertheless, the profit of the QL method is weak. We can state the same previously remarks on the numerical stability and the calculation speed.

5.2 Study on no-Gaussian simulations

We seek here to study the performance of the various methods when the random variables u_t are not Gaussian. In this case, we must use the PML method, and not the ML method like previously. Nevertheless, we give the results obtained by the ML method for benchmark.

Tables 4 and 5: The QL and PML procedures produce the same results. Results obtained by AQL and ML procedures are very close. These two last are more powerful. The adaptive procedure allows to recover the asymptotic effectiveness lost by QL procedure. Let us note that the AQL procedure is numerically more stable and rapid that the ML.

Conclusion

We have described a quasi-likelihood (QL) methodology for estimate the parameters of a model with CH errors. We have shown the convergence and effective properties. This method allows a statistical inference in a general framework, in particular the Gaussian assumption is not useful. The numerical experiments carried out by Monte Carlo simulation show that the procedure suggested is better adapted in this context than the usually procedures. Indeed, in the case of an ARCH process or a process with ARCH errors, procedures QL and PML have the same quality. If the PDF is known the QL is complete by an adaptive procedure allowing to achieve effectiveness of ML estimator. Moreover, our method is easier to compute and more speed in computing time.

Appendix A: Proofs of Section 3

We have $\hat{\alpha}^{(0)} = \left(\frac{1}{T}\sum_{t=1}^{T}\Phi_t'\Phi_t\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\Phi_t'\varepsilon_t^2\right)$ with $\varepsilon_t^2 = \Phi_t\alpha + \eta_t$, the process η_t checks $E(\eta_t/\mathcal{F}_{t-1}) = 0$ and $E(\eta_t^2/\mathcal{F}_{t-1}) = \sigma_t^4 E(u_t^2 - 1)$ thus,

$$(\hat{\alpha}^{(0)} - \alpha) = \left(\frac{1}{T}\sum_{t=1}^{T}\Phi_t'\Phi_t\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\Phi_t'\eta_t\right).$$

As the process (Φ_t, η_t) is strictly stationary and ergodic, according to $E(\varepsilon_t^4) < +\infty$ we have $E(\Phi_t'\Phi_t) < +\infty$ and $E(\Phi_t'\eta_t) = 0$. The ergodic theorem of Hall and Heyde [11] allows to write

$$\frac{1}{T}\sum_{t=1}^{T}\Phi_t'\Phi_t \xrightarrow{a.s.} E(\Phi_t'\Phi_t) \quad \text{and} \quad \frac{1}{T}\sum_{t=1}^{T}\Phi_t'\eta_t \xrightarrow{a.s.} 0$$

and thus $(\hat{\alpha}^{(0)} - \alpha) \xrightarrow{a.s.} 0$.

We have $\sqrt{T}(\hat{\alpha}^{(0)} - \alpha) = \left(\frac{1}{T}\sum_{t=1}^{T}\Phi_t'\Phi_t\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_t'\eta_t\right)$. Like $\eta_t = (u_t^2 - 1)\sigma_t^2$, it comes $\sqrt{T}(\hat{\alpha}^{(0)} - \alpha) = \left(\frac{1}{T}\sum_{t=1}^{T}\Phi_t'\Phi_t\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\Phi_t'(u_t^2 - 1)\sigma_t^2\right)$. Let us consider, $S_t = (u_t^2 - 1)\sigma_t^2\Phi_t'$, then S_t is an martingale increase with

$$\begin{aligned} E(S_t/\mathcal{F}_{t-1}) &= 0 \\ E(S_tS_t'/\mathcal{F}_{t-1}) &= E(u_t^2 - 1)^2\sigma_t^4\Phi_t'\Phi_t. \end{aligned}$$

Moreover

$$\frac{1}{T}\sum_{t=1}^{T}\sigma_t^4\Phi_t'\Phi_t \xrightarrow{a.s.} E(\sigma_t^4\Phi_t'\Phi_t),$$

and according to the central limit theorem for martingales it comes:

$$\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}S_t\right) \xrightarrow{law} \mathcal{N}(0, E(u_t^2 - 1)^2E(\sigma_t^4\Phi_t'\Phi_t)).$$

from where the desired result is obtained

$$\sqrt{T}(\hat{\alpha}^{(0)} - \alpha) \xrightarrow{law} \mathcal{N}(0, B) \text{ with } B = E(u_t^2 - 1)^2E(\Phi_t'\Phi_t)^{-1}E(\sigma_t^4\Phi_t'\Phi_t)E(\Phi_t'\Phi_t)^{-1}.$$

Let us write $(\hat{\alpha}^{(1)} - \alpha) = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\Phi_t'\Phi_t}{\tilde{\sigma}_t^4}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\Phi_t'\eta_t}{\tilde{\sigma}_t^4}\right)$. Like $\tilde{\sigma}_t^2 = \Phi_t\hat{\alpha}^{(0)} = \sigma_t^2 + \Phi_t(\hat{\alpha}^{(0)} - \alpha)$ there is the approximation:

$$\frac{1}{T}\sum_{t=1}^{T}\tilde{\sigma}_t^2 = \frac{1}{T}\sum_{t=1}^{T}\sigma_t^2 + \mathcal{O}_p(T^{-\frac{1}{2}})$$

from where

$$(\hat{\alpha}^{(1)} - \alpha) = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\Phi_t'\Phi_t}{\sigma_t^4}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\Phi_t'\eta_t}{\sigma_t^4}\right) + \mathcal{O}_p(T^{-\frac{1}{2}}).$$

The process σ_t^2 is bounded below by a positive constant and it is strictly stationary and ergodic, moreover $E(\varepsilon_t^4) < +\infty$, i.e. $E(\sigma_t^4) < +\infty$. Therefore the process $(\Phi_t, \eta_t, \sigma_t^{-2})$ is strictly stationary and ergodic. Like $E\left(\frac{\Phi_t'\Phi_t}{\sigma_t^4}\right) < +\infty$ and $E\left(\frac{\Phi_t'\eta_t}{\sigma_t^4}\right) = 0$, according to the ergodic theorem it comes, $\frac{1}{T}\sum_{t=1}^{T}\frac{\Phi_t'\Phi_t}{\sigma_t^4} \xrightarrow{a.s.} E\left(\frac{\Phi_t'\Phi_t}{\sigma_t^4}\right)$ and $\frac{1}{T}\sum_{t=1}^{T}\frac{\Phi_t'\eta_t}{\sigma_t^4} \xrightarrow{a.s.} 0$ from where $(\hat{\alpha}^{(1)} - \alpha) \xrightarrow{a.s.} 0$.

$$\sqrt{T}(\hat{\alpha}^{(1)} - \alpha) = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\Phi_t'\Phi_t}{\sigma_t^4}\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{\Phi_t'\eta_t}{\sigma_t^4}\right) + \mathcal{O}_p(1)$$

$$\sqrt{T}(\hat{\alpha}^{(1)} - \alpha) = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\Phi_t'\Phi_t}{\sigma_t^4}\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{(u_t^2-1)}{\sigma_t^2}\Phi_t'\right) + \mathcal{O}_p(1)$$

Let us consider, $S_t = \frac{(u_t^2-1)}{\sigma_t^2}\Phi_t'$, then S_t is a martingale increase.

$$\begin{aligned} E(S_t/\mathcal{F}_{t-1}) &= 0 \\ E(S_tS_t'/\mathcal{F}_{t-1}) &= E\left(\frac{(u_t^2-1)^2}{\sigma_t^4}\Phi_t'\Phi_t/\mathcal{F}_{t-1}\right) \\ &= E(u_t^2-1)^2\frac{\Phi_t'\Phi_t}{\sigma_t^4} \end{aligned}$$

By the same previous arguments, it is shown that:

$$\sqrt{T}(\hat{\alpha}^{(1)} - \alpha) \longrightarrow \mathcal{N}(0, C^{-1}) \text{ with } C = E(u_t^2-1)^{-2}E\left(\frac{\Phi_t'\Phi_t}{\sigma_t^4}\right).$$

We have

$$(\hat{b}^{(0)} - b) = \left(\frac{1}{T}\sum_{t=1}^{T}x_t'x_t\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}x_t'\varepsilon_t\right)$$

with

$$\frac{1}{T}\sum_{t=1}^{T}x_t'x_t \xrightarrow{a.s.} E(x_t'x_t) < \infty \text{ and } \frac{1}{T}\sum_{t=1}^{T}x_t'\varepsilon_t \xrightarrow{a.s.} E(x_t'\varepsilon_t) = 0$$

because u_t is not correlated with $\{x_s, s \leq t\}$, we deduce immediately that $(\hat{b}^{(0)} - b) \xrightarrow{a.s.} 0$. By using the same previous techniques, it is shown that:

$$\sqrt{T}(\hat{b}^{(0)} - b) \xrightarrow{law} \mathcal{N}(0, P_1) \text{ with } P_1 = E(x_t'x_t)^{-1}E(\sigma_t^2x_t'x_t)E(x_t'x_t)^{-1}.$$

Thus,

$$\hat{\alpha}^{(0)} = \left(\frac{1}{T}\sum_{t=1}^{T}\tilde{\Phi}_t'\tilde{\Phi}_t\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\tilde{\varepsilon}_t^2\tilde{\Phi}_t'\right),$$

with $\tilde{\Phi}_t = (1, \tilde{\varepsilon}_t^2, \cdots, \tilde{\varepsilon}_{t-p}^2)$ and $\tilde{\varepsilon}_t = y_t - x_t\hat{b}^{(0)} = \varepsilon_t - x_t(\hat{b}^{(0)} - b)$. Let us pose $\hat{z}_t = x_t(\hat{b}^{(0)} - b)$, we have the following approximation:

$$\begin{aligned}
\frac{1}{T}\sum_{t=1}^{T}\tilde{\varepsilon}_{t-1}^2 &= \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t-1}^2 - \frac{2}{T}\sum_{t=1}^{T}\varepsilon_{t-1}\hat{z}_t + \frac{1}{T}\sum_{t=1}^{T}\hat{z}_t^2 \\
&= \frac{1}{T}\sum_{t=1}^{T}\varepsilon_{t-1}^2 + \mathcal{O}_p(T^{-\frac{1}{2}}),
\end{aligned}$$

then

$$\begin{aligned}
\sqrt{T}(\hat{\alpha}^{(0)} - \alpha) &= \left(\frac{1}{T}\sum_{t=1}^{T}\tilde{\Phi}_t'\tilde{\Phi}_t\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_t\tilde{\Phi}_t'\right) \\
&= \left(\frac{1}{T}\sum_{t=1}^{T}\Phi_t'\Phi_t + \mathcal{O}_p(T^{-\frac{1}{2}})\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\eta_t\Phi_t' + \mathcal{O}_p(1)\right) \\
&= \left(\frac{1}{T}\sum_{t=1}^{T}\Phi_t'\Phi_t\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2 - 1)\sigma_t^2\Phi_t'\right) + \mathcal{O}_p(1).
\end{aligned}$$

Like we have $\left(\frac{1}{T}\sum_{t=1}^{T}\Phi_t'\Phi_t\right) \xrightarrow{a.s.} E(\Phi_t'\Phi_t)$ and $\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}(u_t^2 - 1)\sigma_t^2\Phi_t'\right) \xrightarrow{a.s.} E(u_t^2 - 1)^2E(\sigma_t^4\Phi_t'\Phi_t)$, we deduce immediatly:

$$\sqrt{T}(\hat{\alpha}^{(0)} - \alpha) \xrightarrow{law} \mathcal{N}(0, B).$$

$$\sqrt{T}(\hat{b}^{(1)} - b) = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{x_t'x_t}{\tilde{\sigma}_t^2}\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{x_t'\tilde{\varepsilon}_t}{\tilde{\sigma}_t^2}\right)$$

$$\sqrt{T}(b^{(1)} - b) = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{x_t'x_t}{\sigma_t^2}\right)^{-1}\left(\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\frac{x_t'\varepsilon_t}{\sigma_t^2}\right) + \mathcal{O}_p(1)$$

We pose $S_t = x_t'\varepsilon_t/\sigma_t^2$, then we can write $(b^{(1)} - b) \xrightarrow{a.s.} 0$ and $\sqrt{T}(b^{(1)} - b) \xrightarrow{law} \mathcal{N}(0, Q_1^{-1})$ with $Q_1 = E\left(\frac{x_t'x_t}{\sigma_t^2}\right)$.

For the α parameter the QGLS estimator is given by

$$\hat{\alpha}^{(1)} = \left(\frac{1}{T}\sum_{t=1}^{T}\frac{\tilde{\Phi}_t'\tilde{\Phi}_t}{\tilde{\sigma}_t^4}\right)^{-1}\left(\frac{1}{T}\sum_{t=1}^{T}\frac{\tilde{\varepsilon}_t^2\tilde{\Phi}_t'}{\tilde{\sigma}_t^4}\right)$$

with $\tilde{\Phi}_t = (1, \tilde{\varepsilon}_t^2, \cdots, \tilde{\varepsilon}_{t-p}^2)$ and $\tilde{\varepsilon}_t = y_t - x_t \hat{b}^{(1)}$. Finally, we have shown that $\sqrt{T}(\hat{\alpha}^{(1)} - \alpha) \stackrel{law}{\longrightarrow} \mathcal{N}(0, C^{-1})$. Note that the asymptotic independence between $\hat{b}^{(1)}$ and $\hat{\alpha}^{(1)}$ comes from a non-correlation between u_t and $u_t^2 - 1$.

Appendix B: Proof of Section 4

By a central limit theorem we have immediately $\Delta_T(\theta_0, f) \longrightarrow \mathcal{N}(0, I(\theta_0, f))$. $\partial l_t(\theta)/\partial\theta$ is a continue difference of martingale with a uniformly limited variance (Bollerslev and Wooldridge [4]).

$$E\left(\Delta_T(\theta_0, f)/\mathcal{F}_{t-1}\right) = 0$$

$$\frac{1}{T}\sum_{t=1}^{T} \frac{\partial l_t(\theta_0)}{\partial\theta} \frac{\partial l_t(\theta_0)}{\partial\theta'} \stackrel{T\to+\infty}{\longrightarrow} I(\theta_0, f)$$

$$E\left(\Delta'_T(\theta_0, f)\Delta_T(\theta_0, f)/\mathcal{F}_{t-1}\right) = I(\theta_0, f)$$

By using the log-likelihood ratio, $L_T(\theta_T) - L_T(\theta_0)$, and the same techniques as in Swensen [20] and Steigerwald [19] we shown

$$L_T(\theta_T) - L_T(\theta_0) - h\Delta_T(\theta_0, f) + \frac{h' I(\theta_0, f) h}{2} \stackrel{a.s.}{\longrightarrow} 0$$

where $h \in \mathbf{R}^m$. Moreover, by a central limit theorem, we have

$$L_T(\theta_T) - L_T(\theta_0) - h\Delta_T(\theta_0, f) \longrightarrow \mathcal{N}\left(-\frac{h' I(\theta_0, f) h}{2}, h' I(\theta_0, f) h\right).$$

Let us $\theta_T = \theta_0 + \dfrac{h}{\sqrt{T}}$ and consider the following expression

$$\begin{aligned}
& \Delta_T(\theta_T, f) - \Delta_T(\theta_0, f) + I(\theta_0, f)\sqrt{T}(\theta_T - \theta_0) \\
= \; & \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\partial l_t(\theta_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right) + I(\theta_0, f)\sqrt{T}(\theta_T - \theta_0) \\
= \; & \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\frac{\partial l_t(\theta_T)}{\partial\theta} - E_0\left(\frac{\partial l_t(\theta_T)}{\partial\theta}/\mathcal{F}_{t-1}\right) - \frac{\partial l_t(\theta_0)}{\partial\theta}\right] + \frac{1}{\sqrt{T}}\sum_{t=1}^{T} E_0\left(\frac{\partial l_t(\theta_T)}{\partial\theta}/\mathcal{F}_{t-1}\right) \\
& + I(\theta_0, f)\sqrt{T}(\theta_T - \theta_0) \\
= \; & A_1 + A_2
\end{aligned}$$

with

$$\begin{aligned}
A_1 &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left[\frac{\partial l_t(\theta_T)}{\partial\theta} - E_0\left(\frac{\partial l_t(\theta_T)}{\partial\theta}/\mathcal{F}_{t-1}\right) - \frac{\partial l_t(\theta_0)}{\partial\theta}\right] \\
A_2 &= \frac{1}{\sqrt{T}}\sum_{t=1}^{T} E_0\left(\frac{\partial l_t(\theta_T)}{\partial\theta}/\mathcal{F}_{t-1}\right) + I(\theta_0, f)\sqrt{T}(\theta_T - \theta_0)
\end{aligned}$$

We must show that $A_1 = o_p(1)$ and $A_2 = o_p(1)$.

$$\begin{aligned}
E\left(\|A_1\|^2\right) &= E\left(\left\|\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\left(\frac{\partial l_t(\theta_T)}{\partial\theta} - E_0\left(\frac{\partial l_t(\theta_T)}{\partial\theta}/\mathcal{F}_{t-1}\right) - \frac{\partial l_t(\theta_0)}{\partial\theta}\right)\right\|^2\right)\\
&\leq \frac{1}{T}\sum_{t=1}^{T} E\left(\left\|\frac{\partial l_t(\theta_T)}{\partial\theta} - E_0\left(\frac{\partial l_t(\theta_T)}{\partial\theta}/\mathcal{F}_{t-1}\right) - \frac{\partial l_t(\theta_0)}{\partial\theta}\right\|^2\right)\\
&\leq \frac{2}{T}\sum_{t=1}^{T} E\left(\left\|\frac{\partial l_t(\theta_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right\|^2\right) + \frac{2}{T}\sum_{t=1}^{T} E\left(\left\|E_0\left(\frac{\partial l_t(\theta_T)}{\partial\theta}/\mathcal{F}_{t-1}\right)\right\|^2\right)\\
&\leq \|A_{11}\| + \|A_{12}\|
\end{aligned}$$

with

$$\|A_{11}\| = \frac{2}{T}\sum_{t=1}^{T} E\left(\left\|\frac{\partial l_t(\theta_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right\|^2\right)$$

$$\|A_{12}\| = \frac{2}{T}\sum_{t=1}^{T} E\left(\left\|E_0\left(\frac{\partial l_t(\theta_T)}{\partial\theta}/\mathcal{F}_{t-1}\right)\right\|^2\right)$$

Lemma 5.1

$$\|A_{11}\| \to 1 \;\; et \;\; \|A_{12}\| \to 1$$

Proof of the lemma:

$$\frac{\partial l_t(\theta)}{\partial\theta} = -\left[\frac{1}{\sigma_t(\theta)}\frac{\partial m_t(\theta)}{\partial\theta}\frac{f'(u_t(\theta))}{f(u_t(\theta))} + \frac{1}{2\sigma_t^2(\theta)}\frac{\partial\sigma_t^2(\theta)}{\partial\theta}\left(1 + u_t(\theta)\frac{f'(u_t(\theta))}{f(u_t(\theta))}\right)\right]$$

let us pose

$$\Psi_1(\theta) = \frac{f'(u_t(\theta))}{f(u_t(\theta))} \;\; \text{et} \;\; \Psi_2(\theta) = 1 + u_t(\theta)\Psi_1(\theta)$$

$$\frac{\partial l_t(\theta_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta} = A_{111} + A_{112}$$

with

$$A_{111} = \frac{1}{\sigma_t(\theta_0)}\frac{\partial m_t(\theta_0)}{\partial\theta}\Psi_1(\theta_0) - \frac{1}{\sigma_t(\theta_T)}\frac{\partial m_t(\theta_T)}{\partial\theta}\Psi_1(\theta_T)$$

$$A_{112} = \frac{1}{2\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\Psi_2(\theta_0) - \frac{1}{2\sigma_t^2(\theta_T)}\frac{\partial\sigma_t^2(\theta_T)}{\partial\theta}\Psi_2(\theta_T)$$

$$\begin{aligned}
A_{11} &= \frac{2}{T}\sum_{t=1}^{T} E\left(\left\|\frac{\partial l_t(\theta_T)}{\partial\theta} - \frac{\partial l_t(\theta_0)}{\partial\theta}\right\|^2\right)\\
&\leq \frac{4}{T}\sum_{t=1}^{T} E\|A_{111}\|^2 + \frac{4}{T}\sum_{t=1}^{T} E\|A_{112}\|^2
\end{aligned}$$

For the CH parameters $A_{111} = 0$, it thus remain to show that $E\|A_{112}\|^2 \to 0$. We have:

$$\begin{aligned}
\|A_{112}\|^2 &= \frac{1}{4}\left\|\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta}\Psi_2(\theta_0) - \frac{1}{\sigma_t^2(\theta_T)}\frac{\partial\sigma_t^2(\theta_T)}{\partial\theta}\Psi_2(\theta_T)\right\|^2 \\
&= \frac{1}{4}\left\|\left(\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta} - \frac{1}{\sigma_t^2(\theta_T)}\frac{\partial\sigma_t^2(\theta_T)}{\partial\theta}\right)\Psi_2(\theta_0) - \frac{1}{\sigma_t^2(\theta_T)}\frac{\partial\sigma_t^2(\theta_T)}{\partial\theta}(\Psi_2(\theta_T) - \Psi_2(\theta_0)\right\|^2 \\
&\leq \frac{1}{2}\left\|\left(\frac{1}{\sigma_t^2(\theta_0)}\frac{\partial\sigma_t^2(\theta_0)}{\partial\theta} - \frac{1}{\sigma_t^2(\theta_T)}\frac{\partial\sigma_t^2(\theta_T)}{\partial\theta}\right)\Psi_2(\theta_0)\right\|^2 \\
&+ \frac{1}{2}\left\|\frac{1}{\sigma_t^2(\theta_T)}\frac{\partial\sigma_t^2(\theta_T)}{\partial\theta}(\Psi_2(\theta_T) - \Psi_2(\theta_0))\right\|^2,
\end{aligned}$$

however

$$\frac{1}{\sigma_t^4(\theta_T)}\|\Psi_2(\theta_0) - \Psi_2(\theta_T)\|^2 \leq \sigma^2\|\Psi_2(\theta_0) - \Psi_2(\theta_T)\|^2$$

and

$$E_0\left(\frac{\partial\sigma_t^2(\theta_T)}{\partial\theta}\frac{\partial\sigma_t^2(\theta_T)}{\partial\theta'}\right) < +\infty$$

moreover $\sigma_t^{-1}(\theta_0)$ is independent of $\sigma_t^{-1}(\theta_T)$ and these quantities depend only on the past.

$$E_0\left[(\sigma_t^{-2}(\theta_T) - \sigma_t^{-2}(\theta_0))\Psi_2(\theta_0)\right]^2 = E_0\left[(\sigma_t^{-2}(\theta_T) - \sigma_t^{-2}(\theta_0))\right]^2 I_2$$

however like

$$\max_{1\leq t\leq T} E_0\left[(\sigma_t^{-2}(\theta_T) - \sigma_t^{-2}(\theta_0))\right]^2 \to 0$$

from where it is deduced that $E\|A_{112}\|^2 \to 0$.

For the conditional mean parameter we have

$$\frac{1}{\sigma_t^2(\theta_T)}\|\Psi_1(\theta_0) - \Psi_1(\theta_T)\|^2 \leq \sigma\|\Psi_1(\theta_0) - \Psi_1(\theta_T)\|^2,$$

Ψ_1 is independent of $\sigma_t^{-1}(\theta_0)$ and $\sigma_t^{-1}(\theta_T)$ bus these quantities depend only on the past. In the same way we have

$$E_0\left[(\sigma_t^{-1}(\theta_T) - \sigma_t^{-1}(\theta_0))\Psi_1(\theta_0)\right]^2 = E_0\left[(\sigma_t^{-1}(\theta_T) - \sigma_t^{-1}(\theta_0))\right]^2 I_1$$

however like

$$\max_{1\leq t\leq T} E_0\left[(\sigma_t^{-1}(\theta_T) - \sigma_t^{-1}(\theta_0))\right]^2 \to 0$$

from where we deduced that $A_{11} = o_p(1)$.

To show that $A_{12} = o_p(1)$ we use the same technique while noticing that:

$$E_0E(\Psi_1(\theta_T)/\mathcal{F}_{t-1}) \longrightarrow 0.$$

We show other relations by similar reasoning.

Appendix C: Simulations tables

Parameters α_0 α_1 α_2 α_3	Size	Estimator	Mean α_0	Mean α_1	Mean α_2	Mean α_3	Std α_0	Std α_1	Std α_2	Std α_3	No admissible
0.2 0.3 0.2 0.3	T=300	OLS	0.4064	0.2072	0.1413	0.2088	0.2775	0.1057	0.1019	0.1112	23
		QGLS	0.2335	0.2923	0.1750	0.2900	0.0550	0.0863	0.0809	0.1074	9
		QL	0.2117	0.3027	0.1801	0.2994	0.0475	0.0980	0.0846	0.1065	9
		ML	0.2118	0.3027	0.1803	0.2995	0.0475	0.0979	0.0846	0.1064	10
	T=500	OLS	0.3715	0.1911	0.1801	0.1981	0.1186	0.0855	0.0962	0.1091	20
		QGLS	0.2287	0.2862	0.1983	0.2854	0.0609	0.0728	0.0752	0.1016	6
		QL	0.2050	0.2925	0.1933	0.2974	0.0583	0.0692	0.0672	0.0986	5
		ML	0.2050	0.2925	0.1933	0.2974	0.0583	0.0692	0.0672	0.0986	7
0.05 0.25 0.2 0.5	T=300	OLS	0.0715	0.1911	0.1873	0.3518	0.0647	0.1251	0.1135	0.1316	55
		QGLS	0.0613	0.2064	0.1887	0.4264	0.0149	0.0888	0.0729	0.1249	40
		QL	0.0576	0.2152	0.1911	0.4278	0.0136	0.0857	0.0735	0.1240	38
		ML	0.0577	0.2152	0.1911	0.4278	0.0136	0.0857	0.0735	0.1240	50
	T=500	OLS	0.1071	0.1970	0.1576	0.3932	0.1879	0.1030	0.0851	0.1292	53
		QGLS	0.0579	0.2272	0.1777	0.4545	0.0126	0.0690	0.0535	0.0813	35
		QL	0.0549	0.2321	0.1810	0.4585	0.0104	0.0710	0.0539	0.0849	33
		ML	0.0549	0.2321	0.1810	0.4585	0.0104	0.0710	0.0539	0.0849	47

Table 1: Gaussian ARCH(3) Model

Parameters α_0 α_1	Size	Estimator	Mean α_0	Mean α_1	Std α_0	Std α_1	No admissible
0.2 0.8	T=300	OLS	0.1917	0.7682	0.1641	0.1559	-
		QGLS	0.2189	0.6947	0.0330	0.1267	-
		QL	0.2185	0.6925	0.0356	0.1291	-
		ML	0.2185	0.6925	0.0356	0.1291	2
	T=500	OLS	0.2170	0.7334	0.2447	0.2404	-
		QGLS	0.2067	0.7239	0.0230	0.0949	-
		QL	0.2025	0.7335	0.0232	0.0987	-
		ML	0.2025	0.7335	0.0232	0.0987	-
0.05 0.95	T=300	OLS	0.0768	0.7583	0.1334	1.1551	35
		QGLS	0.0542	0.8685	0.0112	0.1020	20
		QL	0.0524	0.8713	0.0083	0.0991	22
		ML	0.0524	0.8713	0.0083	0.0991	39
	T=500	OLS	0.0643	0.8088	0.1267	0.1345	22
		QGLS	0.0510	0.8585	0.0066	0.1060	18
		QL	0.0509	0.8593	0.0063	0.1060	17
		ML	0.0509	0.8593	0.0063	0.1060	30

Table 2: Gaussian ARCH(1) Model

Parameters b α_0 α_1	Size	Estimator	Mean b	Mean α_0	Mean α_1	Std b	Std α_0	Std α_1	No convergence
2.5 0.2 0.8	T=300	QGLS	2.4997	0.2695	0.6660	0.0356	0.0543	0.1269	13
		QL	2.5016	0.2146	0.7017	0.0339	0.0294	0.1291	7
		AQL	2.5014	0.2154	0.6935	0.0246	0.0309	0.1329	7
		ML	2.4985	0.2244	0.6894	0.0314	0.0563	0.1328	31
	T=500	QGLS	2.5020	0.2493	0.7200	0.0277	0.0378	0.0821	7
		QL	2.5022	0.1992	0.7571	0.0274	0.0234	0.0962	-
		AQL	2.5044	0.2003	0.7480	0.0254	0.0235	0.0964	-
		ML	2.5040	0.2011	0.7533	0.0249	0.0237	0.0960	24
2.5 0.05 0.95	T=300	QGLS	2.5039	0.0757	0.7505	0.0157	0.0197	0.1131	24
		QL	2.5035	0.0531	0.8373	0.0157	0.0067	0.1016	33
		AQL	2.5022	0.0533	0.8300	0.0165	0.0071	0.1002	29
		ML	2.5026	0.0562	0.8223	0.0174	0.0147	0.1048	61
	T=500	QGLS	2.4990	0.0747	0.8284	0.0122	0.0265	0.1530	32
		QL	2.4990	0.0526	0.8505	0.0120	0.0057	0.0934	30
		AQL	2.4961	0.0530	0.8420	0.0111	0.0059	0.0930	30
		ML	2.4967	0.0561	0.8412	0.0113	0.0064	0.0977	73

Table 3: Regression Model with Gaussian ARCH(1) errors

Parameters α_0 α_1	Size	Estimator	Mean α_0	Mean α_1	Std α_0	Std α_1	No admissible
0.2 0.8	T=300	QGLS	0.2219	0.5968	0.0521	0.1865	21
		QL	0.2132	0.6234	0.0544	0.1980	30
		PML	0.2132	0.6234	0.0544	0.1979	31
		AQL	0.2189	0.6935	0.0366	0.1311	2
		ML	0.2181	0.6927	0.0354	0.1323	5
	T=500	QGLS	0.2127	0.6698	0.0515	0.1792	16
		QL	0.2106	0.6839	0.0464	0.1734	18
		PML	0.2106	0.6839	0.0464	0.1734	19
		AQL	0.2037	0.7345	0.0251	0.0998	-
		ML	0.2029	0.7318	0.0242	0.1013	-
0.05 0.95	T=300	QGLS	0.0590	0.6640	0.0158	0.1670	24
		QL	0.0570	0.6919	0.0152	0.1650	29
		PML	0.0570	0.6915	0.0152	0.1662	30
		AQL	0.0528	0.8763	0.0088	0.0985	22
		ML	0.0526	0.8747	0.0085	0.0992	37
	T=500	QGLS	0.0542	0.7231	0.0130	0.1552	27
		QL	0.0534	0.7389	0.0122	0.1541	34
		PML	0.0534	0.7389	0.0122	0.1541	35
		AQL	0.0513	0.8623	0.0066	0.1043	16
		ML	0.0511	0.8601	0.0065	0.1052	31

Table 4: ARCH(1) Model with $\mathcal{G}(1)$ law, kurtosis=9, skewness=2

Parameters b α_0 α_1	Size	Estimator	Mean b	Mean α_0	Mean α_1	Std b	Std α_0	Std α_1	No convergence
2.5 0.2 0.8	T=300	QGLS	2.4975	0.2680	0.6196	0.0274	0.0698	0.1958	12
		QL	2.4974	0.2163	0.6594	0.0262	0.0578	0.1833	20
		AQL	2.5008	0.2062	0.6415	0.0281	0.0560	0.1798	23
		PML	2.5037	0.2164	0.6631	0.0295	0.0562	0.1822	21
		ML	2.5007	0.2054	0.6531	0.0280	0.0552	0.1819	18
	T=500	QGLS	2.5016	0.2673	0.6828	0.0294	0.0502	0.2884	19
		QL	2.5019	0.2074	0.6955	0.0284	0.0446	0.1987	20
		AQL	2.5021	0.2072	0.6863	0.0269	0.0450	0.1986	18
		PML	2.5024	0.2094	0.6848	0.0283	0.0457	0.2004	21
		ML	2.5019	0.2091	0.6868	0.0268	0.0452	0.1991	20
2.5 0.05 0.95	T=300	QGLS	2.5006	0.0707	0.6369	0.0186	0.0192	0.2032	29
		QL	2.5010	0.0545	0.7015	0.0170	0.0127	0.2133	39
		AQL	2.5002	0.0543	0.6867	0.0145	0.0129	0.1998	34
		PML	2.4999	0.0562	0.6968	0.0159	0.0153	0.1821	43
		ML	2.5001	0.0544	0.6878	0.0142	0.0132	0.1991	39
	T=500	QGLS	2.5028	0.0798	0.7619	0.0134	0.0541	0.3829	29
		QL	2.5029	0.0561	0.7578	0.0129	0.0116	0.1704	37
		AQL	2.5010	0.0560	0.7460	0.0119	0.0115	0.1614	34
		PML	2.5020	0.0683	0.7311	0.0126	0.0825	0.1546	37
		ML	2.5009	0.0564	0.7478	0.0117	0.0117	0.1691	36

Table 5: Regression Model with $\mathcal{G}(1)$-ARCH(1) errors, kurtosis=9, skewness=2

References

[1] A.K. Bera and M.L. Higgins. ARCH Models : Properties, Estimation and Testing. *Journal of Economic Surveys*, 7(4):305–366, 1993.

[2] E.K. Bernt, B.H. Hall, L.R.E. Hall, and J.A. Hausman. Estimation and inference in non linear structural models. *Annals of Economic and Social Measurement*, 3:653–666, 1974.

[3] T. Bollerslev, R.Y. Chou, and K.F. Kroner. ARCH modeling in finance; A review of the theory and empirical evidence. *Journal of Econometrics*, 52:5–59, 1992.

[4] T. Bollerslev and J. Wooldridge. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews*, 1992.

[5] R.F. Engle. Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica*, 54:987–1008, 1982.

[6] R.F. Engle and G Gonzàlez-Rivera. Semi-parametric ARCH models. *Journal of Business and Economic Statistics*, 9:345–359, 1991.

[7] V.P. Godambe and C.C. Heyde. Quasi-likelihood and optimal estimation. *Internat. Statist. Rev.*, 55:231–244, 1987.

[8] G. González-Rivera and J.S. Racine. Maximum likelihood estimation and testing strategies in garch models. Technical report, University of California, Riverside (UCR), 1995.

[9] C. Gourieroux and A. Monfort. *Statistique et modèles économétriques*, volume 1 et 2. Economica, 2 edition, 1996.

[10] C. Gourieroux, A. Monfort, and A. Trognon. Pseudo maximum likelihood methods: Theory. *Econometrica*, 52:681–700, 1984.

[11] P. Hall and C.C. Heyde. *Martingale limit theory and its applications.* Academic Press, San Diego, 1980.

[12] C.C. Heyde. *Quasi-Likelihood And Its Application.* Springer Series in Statistics. Springer, 1997.

[13] N.L. Johnson and S. Kotz. *Continuous Univariate Distributions.* Wiley, 1970.

[14] S. Lee and B. Hansen. Asymptotic theory for the GARCH(1,1) quasi-maximum likelihood estimator. *Econometric Theory*, 10:29–52, 1994.

[15] R. Lumsdaine. Asymptotic properties of the quasi-maximum likelihood estimator in GARCH(1,1) and IGARCH(1,1) models. Technical report, Department of Economics, Princeton University, 1992.

[16] P. McCullagh. Quasi-likelihood functions. *The Annals of Statistics*, 11(1):59–67, 1983.

[17] P. McCullagh and J.A. Nelder. *Generalized linear models.* Chapman and Hall, 2nd edition, 1989.

[18] S.G. Pantula. Estimation of autoregressive models with ARCH errors. *Sankhyã*, B(50):119–138, 1988.

[19] D. Steigerwald. Efficient estimation of models with conditional heteroscedasticity. Technical report, Department of Economics, University of California, Santa Barbara, 1994.

[20] A.R. Swensen. The asymptotic distribution of the likelihood ratio for autoregressive time series with a regression trend. *Journal of Multivariate Analysis*, 16:54–70, 1985.

[21] R.W.M. Wedderburn. Quasi-likelihood functions, generalized linear models and the Gauss-Newton method. *Biometrika*, 61:439–447, 1974.

[22] A.A. Weiss. Asymptotic theory for ARCH models: Estimation and testing. *Econometric Theory*, 2:107–113, 1986.

[23] H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50:483–499, 1982.

[24] J. Wooldridge. *Estimation and inference for dependent process.* Handbook of Econometrics, 1995.

Table of Contents

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LÈS NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l'Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
http://www.inria.fr
ISSN 0249-6399