

Article

# Selection Criteria in Regime Switching Conditional Volatility Models

Thomas Chuffart

Aix-Marseille University (Aix Marseille School of Economics), CNRS & EHESS, Marseille, 13002, France; E-Mail: thomas.chuffart@univ-amu.fr; Tel: +33-4-91140727

Academic Editor: Kerry Patterson

Received: 13 January 2015 / Accepted: 28 April 2015 / Published: 11 May 2015

---

**Abstract:** A large number of nonlinear conditional heteroskedastic models have been proposed in the literature. Model selection is crucial to any statistical data analysis. In this article, we investigate whether the most commonly used selection criteria lead to choice of the right specification in a regime switching framework. We focus on two types of models: the Logistic Smooth Transition GARCH and the Markov-Switching GARCH models. Simulation experiments reveal that information criteria and loss functions can lead to misspecification; BIC sometimes indicates the wrong regime switching framework. Depending on the Data Generating Process used in the experiments, great care is needed when choosing a criterion.

**Keywords:** conditional volatility; model selection; GARCH; regime switching

**JEL classifications:** C15, C22, C52

---

## 1. Introduction

Ever since Engle [1] developed the Autoregressive Conditional Heteroskedasticity (ARCH) models which provide a fruitful framework to analyze volatility and financial time series, this has been a major research focus in financial econometrics. Bollerslev [2] subsequently proposed Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models, where volatility is a linear function of past volatility and squared residual past shocks. These models are of the form  $\epsilon_t = \eta_t \sqrt{h_t}$  where  $h_t$  is a positive process (volatility) and  $\eta_t$  an identically and independently distributed random variable with zero mean and unit variance. Although GARCH models are attractive, copious empirical evidence in

the econometric literature argues against their suitability. For example, these models do not adequately fit the data over a long period of time. Lamoureux and Lastrapes [3] show that if structural changes are not considered, upward GARCH estimates of persistence in variance may be biased. Moreover, while the squares of a GARCH(1,1) process follow the dynamics of an ARMA process with an autocorrelation function (ACF) which goes to zero very fast, the sample ACFs of the squares tend to stabilize around a positive value for larger lags. This is known as long range dependence in volatility. To circumvent these problems, practitioners can use other types of model, such as regime switching models. Several regime switching behaviors exist: in this paper we focus on regime switches driven by a hidden Markov Chain and regime switches driven by a transition function. Markov Switching GARCH (MS-GARCH) models belong to the first class. They have stochastic regime switches, and volatility can therefore take different forms depending on probabilities. Asymmetric GARCH models belong to the second class. The volatility has deterministic regime switches: it depends on a transition variable through a particular transition function, for example, a logistic smooth function which gives the Logistic Smooth Transition GARCH (LST-GARCH) model.

Practitioners studying empirical data need to choose a specific model to estimate conditional volatility. The wrong model will lead to incorrect interpretation of data. For example, selecting the wrong type of regime switching could cause a structural change to be confused with asymmetry. A well-known citation from Box [4] says that “*Models, of course, are never true but fortunately it is only necessary that they be useful*”. Thus, models are just an approximation of the true Data Generating Process (DGP). The challenge for data analysts is to select the best model, the one closest to the DGP. All the inferences and evaluations of real life data depend on accurate specifications yielding good in-sample forecasting results and more accurate interpretation of the economic world. This is known as goodness of fit. In practice, a set of plausible models is selected and narrowed down according to one of a number of criteria. The most popular are Information Criteria (IC), loss functions, and the  $R^2$  from a regression. In time series analysis, IC and in-sample forecasts evaluated with loss functions are very often used. In contrast with out-sample forecasts, where in a lot of cases simpler models provide better results, in-sample forecasts should be better when the closest model is estimated.

In this article, we seek to identify whether certain criteria are more likely to lead to a good choice of regime switching type. To do so, we perform Monte Carlo experiments. We simulate data according to DGPs. We focus on two models: MS-GARCH and LST-GARCH. The estimation of LST-GARCH has been widely discussed in the literature; methods proposed include Quasi Maximum Likelihood (QML) estimation ([5,6]), Generalized Method of Moments ([7]) and Bayesian estimation ([8,9]). All these methods have advantages and drawbacks. We focus on the QML method here, since it is very commonly used in empirical applications. To our knowledge, the consistency and the asymptotic normality of LST-GARCH QML estimation has not been established yet<sup>1</sup>. There appear to be no findings on MS-GARCH models to date, nor on the asymptotic distribution of the Maximum Likelihood (ML) estimation. For example, Augustyniak [12] conducts an experiment which shows that Gray’s method does not generate consistent estimates for the path-dependent MS-GARCH model. These

---

<sup>1</sup> Some empirical studies have shown that the QML estimation of smooth transition models can cause problems in interpretation. See Chan *et al.* [10] and Novella [11].

models suffer from a lack of theoretical grounding, and our aim here is to investigate the consequences on goodness-of-fit. Our experimental task is to estimate the data generated, using a wide variety of specifications. Finally, we apply various selection criteria to these estimations to see which model is chosen by the criteria. This method was employed in recent studies to highlight pitfalls in smooth transition models ([11]).

Our results raise interesting questions, revealing that selection criteria lead to misspecification in some cases. Does this mean that the criteria are not suitable for regime switching conditional volatility models? Probably not: we argue that if these criteria lead the choice of the wrong specification it is more because these regime switching models are difficult to estimate using the QML estimation, especially when the regimes are poorly identified in MS-GARCH models.

The remainder of this article is organized as follow out as follows. In Section 2, we briefly explain the different types of models and the selection criteria we explore. We present our simulation experiment framework in Section 3. Section 4 presents the results and the discussion. Section 5 contains concluding remarks.

## 2. Theory: Models and Selection Criteria

### 2.1. Models

In this section we describe the three classes of GARCH-type models that interest us here, explaining why we focus on these models. First, we describe the class of univariate GARCH models. Secondly, we present some asymmetric GARCH-type models: the LST-GARCH of Hagerud [6] and Gonzalez-Rivera [13], the Glosten-Jagannathan-Runkle (GJR) GARCH and the Exponential GARCH (EGARCH). Then, MS-GARCH-type models are introduced.

#### 2.1.1. Univariate GARCH Model

The univariate GARCH model was introduced by Bollerslev [2] and many other followed. The GARCH model is a benchmark in volatility modeling since it can capture the main stylized facts of a financial time series that exhibit time-varying volatility clustering. A process  $(\epsilon_t)$  is called a GARCH(p,q) process if for  $t = 1, \dots, T$ , with  $T$  the sample size and

$$E(\epsilon_t | \epsilon_s, s < t) = 0, t = 1, \dots, T \quad (1)$$

$$\epsilon_t = \eta_t \sqrt{h_t}, \eta_t \sim IID(0, 1) \quad (2)$$

with  $\eta_t$  an identically and independent distributed random variable with zero mean and unit variance and if there exist  $\omega, \alpha_j, j = 1, \dots, q$  and  $\beta_i, i = 1, \dots, p$  such that

$$h_t = Var(\epsilon_t | \epsilon_u, u < t) = \omega + \sum_{j=1}^q \alpha_j \epsilon_{t-j}^2 + \sum_{i=1}^p \beta_i h_{t-i} \quad (3)$$

GARCH-type models are generally estimated with the Gaussian Quasi Maximum Likelihood (QML) method. A QML estimation of the vector parameters  $\theta_0 = (\omega_0, \alpha_{01}, \dots, \alpha_{0q}, \beta_{01}, \dots, \beta_{0p})'$  is defined as any solution  $\hat{\theta}^{QML}$  of

$$\hat{\theta}^{QML} = \arg \max \mathcal{L} = n^{-1} \sum_{t=1}^T \ell_t \quad (4)$$

with

$$\ell_t = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log h_t - \frac{1}{2} \frac{\epsilon_t^2}{h_t} \quad (5)$$

Gaussian QML estimation assumes that  $\epsilon_t$  are normally distributed. However, if this assumption is not verified,  $\hat{\theta}^{QML}$  is still consistent but can be inefficient, as shown in [14] and [15].

### 2.1.2. Asymmetric Volatility Models

This kind of model belongs to the class of “asymmetric” or “leverage” volatility models. They were introduced in response to empirical evidence that the increase in volatility is larger when returns are negative than when they are positive<sup>2</sup>. This characteristic is known as the “leverage-effect”. These models can be seen as regime switching GARCH models. They have different specifications according to the sign of the past shocks. We consider in this article three different asymmetric GARCH-type models: the LST-GARCH [6], the GJR-GARCH [20] and the EGARCH [21]. A process  $(\epsilon_t)$  is called an LST-GARCH(p,q) process if there exist  $\omega$ ,  $\alpha_{1j}$ ,  $\alpha_{2j}$ ,  $j = 1, \dots, q$ ,  $\beta_i$ ,  $i = 1, \dots, p$ , and  $\gamma$  such that

$$h_t = \omega + \sum_{j=1}^q [\alpha_{1j} + \alpha_{2j} F(\epsilon_{t-j})] \epsilon_{t-j}^2 + \sum_{i=1}^p \beta_i h_{t-i} \quad (6)$$

where

$$F(\epsilon_{t-j}) = (1 + \exp(\gamma \epsilon_{t-j}))^{-1} - \frac{1}{2} \quad (7)$$

with  $\gamma$ , the so-called transition parameter. The LST-GARCH process can be seen as a generalization of the well-known GJR-GARCH. A process  $(\epsilon_t)$  is called a GJR-GARCH(p,q) process if there exist  $\omega$ ,  $\alpha_{1j}$ ,  $\alpha_{2j}$ ,  $j = 1, \dots, q$  and  $\beta_i$ ,  $i = 1, \dots, p$  such that

$$h_t = \omega + \sum_{j=1}^q [\alpha_{1j} + \alpha_{2j} \mathbb{1}(\epsilon_{t-1} > 0)] \epsilon_{t-j}^2 + \sum_{i=1}^p \beta_i h_{t-i} \quad (8)$$

where  $\mathbb{1}(\cdot)$  is the indicator function. When  $\gamma \rightarrow +\infty$ , the logistic function becomes a double step function like the indicator function in the GJR-GARCH. In terms of regime, because the logistic function is continuous, Gonzalez-Rivera [13] talks about a “continuum” of regimes where the probability of regime switching is one. However, the term “regimes” does not have the same meaning as in the MS-GARCH models. Conditions for a stationary positive process are given in Hagerud [6] and Gonzalez-Rivera [13]. There is empirical evidence of how difficult it is to estimate the transition

<sup>2</sup> See [16–19].

parameter<sup>3</sup>. We will therefore try to determine whether this inefficient selection criteria. Finally, we consider the popular EGARCH model by Nelson [21]. A process  $(\epsilon_t)$  is called an EGARCH(p,q) process if there exist  $\omega$ ,  $\alpha_{1j}$ ,  $\alpha_{2j}$ ,  $j = 1, \dots, q$  and  $\beta_i$ ,  $i = 1, \dots, p$

$$\log(h_t) = \omega + \sum_{j=1}^q g_j(\eta_{t-j}) + \sum_{i=1}^p \beta_i \log(h_{t-i}) \quad (9)$$

where

$$g_j(\eta_{t-j}) = \alpha_{1j}\eta_{t-j} + \alpha_{2j}(|\eta_{t-j}| + E|\eta_t|) \quad (10)$$

The function  $g_j$  depends on the magnitude and sign of  $\eta_t$ . This makes it possible to respond asymmetrically to positive and negative values of the error term. As with the GARCH model, estimation is performed by maximizing the log-likelihood function given by  $\mathcal{L} = \sum_{t=1}^T \ell_t$  with  $\ell_t$  represented by Equation (5) assuming the normality of the error term.

### 2.1.3. MS-GARCH Models

These models give rise to a conditional mixture distribution. They allow time-varying skewness contrary to the traditional GARCH-type models: asymmetry exists in the conditional return distribution but this asymmetry is time-varying ([22]). Their main difference from the previous model is the following, for  $t = 1, \dots, T$ :

$$r_t = \epsilon_t$$

with

$$\epsilon_t = \eta_t \sqrt{h_t(\Delta_t)}$$

and  $\eta_t$  an identically and independently distributed random variable with zero mean and unit variance.  $\Delta_t$  is a variable which indicates the state of the world at time  $t$ . We will focus on MS-GARCH processes:  $\Delta_t$  follows a Markov chain with finite state spaces  $S = 1, \dots, k$ , and a transition matrix  $P$ . However, other models exist where transition probabilities have a different definition. In contrast to the LST-GARCH process, the probability of switching from one regime to another is no longer equal to one but depends on the transition matrix  $P$ , given by

$$P = \begin{pmatrix} p_{11} & \dots & p_{k1} \\ \vdots & \dots & \vdots \\ p_{1k} & \dots & p_{kk} \end{pmatrix}$$

with  $p_{ij} = p(\Delta_t = j | \Delta_{t-1} = i)$  the probability of being in state  $j$  at time  $t$ , given of being in state  $i$  at time  $t - 1$ . In this sense, this type of regime switch is endogenous. A process  $(\epsilon_t)$  is called a MS(k)-GARCH(p,q) process if

$$\epsilon_t = \eta_t \sqrt{h_t(\Delta_t)}, \quad \eta_t \sim IID(0, 1) \quad (11)$$

<sup>3</sup> See [11] for example.

and there exist  $\omega(\Delta_t)$ ,  $\alpha_i(\Delta_t)$ ,  $i = 1, \dots, q$ ,  $\beta_l(\Delta_t)$ ,  $l = 1, \dots, p$  and  $j = 1, \dots, k$  such that

$$h_t(\Delta_t) = \omega(\Delta_t) + \sum_{i=1}^q \alpha_i(\Delta_t) \epsilon_{t-q}^2 + \sum_{l=1}^p \beta_l(\Delta_t) h_{t-l} \quad (12)$$

This model cannot be estimated by QML. The calculation of the likelihood function for a sample of  $T$  observations is infeasible because it requires the integration of  $k^T$  possible regime paths ([23] and [24]). To circumvent the path dependence problem, Gray [5] introduces an MS-GARCH model under the hypothesis that the conditional variance at any regime depends on the expectation of previous conditional variance. He proposes to replace  $h_{t-1}$  by the conditional variance of the error term  $\epsilon_{t-1}$  given the information up to  $t-2$ . Klaassen [25] enlarges the information set to  $t-1$  by conditioning the expectation of previous conditional variances on all available observations and also on the current regime:

$$h_t(\Delta_t) = \omega(\Delta_t) + \alpha(\Delta_t) \epsilon_{t-1}^2 + \beta(\Delta_t) \sum_{i=1}^k p(\Delta_{t-1} = i | \Omega_{t-1}, \Delta_t = j) h_{i,t-1} \quad (13)$$

with  $j = 1, \dots, k$  and  $\Omega_t$  is the information set of the process (*i.e.*, the return history up to date  $t-1$ ). The model of Klaassen is the second model of interest. The model of Haas *et al.* [26] contrasts with this approach where each specific conditional variance depends only on its own lag,

$$h_t(\Delta_t) = \omega(\Delta_t) + \alpha(\Delta_t) \epsilon_{t-1}^2 + \beta(\Delta_t) h_{t-1}(\Delta_t) \quad (14)$$

This model can be rewritten in matrix form:

$$\mathbf{h}_t = \boldsymbol{\omega} + \boldsymbol{\alpha} \epsilon_{t-1}^2 + \boldsymbol{\beta} \mathbf{h}_{t-1}$$

where  $\boldsymbol{\omega} = [\omega_1, \omega_2, \dots, \omega_j]'$ ,  $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \dots, \alpha_j]'$  and  $\boldsymbol{\beta} = \text{diag}(\beta_1, \beta_2, \dots, \beta_j)$ .  $\mathbf{h}_t$  is thus a vector of  $k \times 1$  components. In this specification, every regime can be represented as an  $ARCH(\infty)$ , which is the direct generalization of the single-regime GARCH model. This specification permits practitioners to interpret the coefficients in the same way as in the single regime framework. These two MS-GARCH models<sup>4</sup> can easily be estimated by Maximum Likelihood (ML) estimation following the work of Hamilton [28]. An ML estimation of  $\theta_0$ , with the vector parameters  $\theta_0 = (\boldsymbol{\omega}_0, \boldsymbol{\alpha}_0, \boldsymbol{\beta}_0)'$  to be estimated is defined as any solution of  $\hat{\theta}^{ML}$  of

$$\hat{\theta}^{ML} = \arg \max \mathcal{L} = \sum_{t=1}^T \log f(\epsilon_t | \Omega_{t-1})$$

where  $f(\epsilon_t | \Omega_{t-1})$  is the conditional density of  $\epsilon_t$  given the process up to time  $t$ . This density is the sum of conditional regime densities weighted by the conditional regime probabilities  $Pr(\Delta_t = j | \Omega_{t-1}, \theta)$ .

<sup>4</sup> There are a number of expansions of these two MS-GARCH processes. For example, Gallo and Otrento [27] introduce asymmetric effects in each regime variance.

These probabilities are obtained by using the recursive scheme proposed by [29]. Moreover, in-sample forecasts of volatility are computed using these predicted probabilities:

$$\hat{h}_t = \sum_{j=1}^k Pr(\Delta_t = j | \Omega_{t-1}, \hat{\theta}^{ML}) h_{t,j} \quad (15)$$

Finally, an assumption on conditional regime densities has to be made. In this paper, we investigate a case where the conditional regime densities follow a Gaussian density. As noted by [25,30,31], if regimes are not normal but leptokurtic, the use of within-regime normality affects identification of the regime process. It implies, for example, that ML estimation based on Gaussian components does not provide a consistent estimator. Hereafter, for convenience, we call the processes MSG-GARCH-K and MSG-GARCH-H respectively, for Klaassen [25] and Haas *et al.* [26], assuming Gaussian conditional regime densities. Moreover, we call the processes MST-GARCH-K and MST-GARCH-H respectively, for Klaassen [25] and Haas *et al.* [26], assuming Student conditional regime densities.

## 2.2. Selection Criteria: Information Criteria and Loss Functions

In empirical analysis, model selection is based on different methods. A statistical specification test comparing LST-GARCH and MS-GARCH processes does not exist yet<sup>5</sup>. The remaining way of choosing a specification for a model is to use the selection criteria. We focus on IC and in-sample forecasts. In contrast to hypothesis testing, they can be used to compare two models which have different specifications and which are not necessarily nested. According to Akaike [33], a model should be selected on the basis of good results when it is used for prediction. He proposed the well-known AIC to evaluate models in terms of Kullback-Leibler (KL) information ([34]),

$$AIC = 2m - 2 \log(\mathcal{L}(\hat{\theta}|X))$$

where  $\mathcal{L}(\hat{\theta}|X)$  represents the value of the likelihood function in  $\hat{\theta}$ , the vector of the estimated parameters, given the observed data and  $m$  the number of parameters. In general linear models, AIC tends not to perform well in small samples but the criterion tends to select the right model in large samples as shown in [35]. BIC has a similar form,

$$BIC = m \log(T) - 2 \log(\mathcal{L}(\hat{\theta}|X))$$

with  $T$  the sample size. AIC imposes less of a penalty on the number of parameters than does BIC. In contrast to AIC, BIC is reputed to perform poorly in small samples in the context of general linear models.

In-sample forecasting can also be used to compare the real volatility with the predicted volatility. Loss functions measure the difference, and the model with the lowest loss is selected. However, volatility

<sup>5</sup> Hu and Shin [32] introduced a test procedure which tests the null hypothesis of a GARCH process against an MS-GARCH process.



is a latent variable: in practice we do not observe real volatility, so we use proxies to compute it. Hansen and Lunde [36] show that the sufficient condition for a loss function to be robust is that  $\frac{\partial L(\sigma^2, h)}{\partial (\sigma^2)^2}$  does not depend on  $h$ , with  $\hat{\sigma}^2$  the proxy and  $h$  the predicted volatility. Patton [37] derives necessary sufficient conditions. Moreover, he provides a new class of loss functions that guarantee the consistency of the ranking (asymptotically) when the unbiased volatility proxy is used instead of true volatility. The following class requires the assumption that the volatility proxy is unbiased. Loss functions are defined as

$$R(\hat{\sigma}, h, b) = \begin{cases} \frac{1}{(b+1)(b+2)}(\hat{\sigma}^{2b+4} - h^{b+2}) - \frac{1}{b+1}(h^{b+1}(\hat{\sigma}^2 - h)) & b \notin \{-1, -2\} \\ \hat{\sigma}^2 \log \frac{\hat{\sigma}^2}{h} - (\hat{\sigma}^2 - h) & b = -1 \\ \frac{\hat{\sigma}^2}{h} - \log \frac{\hat{\sigma}^2}{h} - 1 & b = -2 \end{cases} \quad (16)$$

with  $b$  a scalar parameter. In this article, we focus on three loss functions: Mean Squared Error (MSE), Quasi Likelihood (QLIKE) and Mean Absolute Error (MAE). MSE and QLIKE are special cases of Equation (16) when  $b = 0$  and  $b = -2$  respectively whereas MAE is not a robust loss function. For  $t = 1, \dots, T$  they are each equal to

$$\begin{aligned} MSE : L(\hat{\sigma}_t, h_t) &= \frac{1}{T} \sum_{t=1}^T (\hat{\sigma}_t^2 - h_t)^2 \\ QLIKE : L(\hat{\sigma}_t, h_t) &= \frac{1}{T} \sum_{t=1}^T \left( \frac{\hat{\sigma}_t^2}{h_t} - \log \frac{\hat{\sigma}_t^2}{h_t} - 1 \right) \\ MAE : L(\hat{\sigma}_t, h_t) &= \frac{1}{T} \sum_{t=1}^T |\hat{\sigma}_t^2 - h_t| \end{aligned}$$

One important difference between MSE and QLIKE is that the former treats positive and negative errors equally, while the latter imposes a larger penalty when the forecast underestimates the realized volatility.

### 3. Design of the Experiments

The objective of our study was to perform Monte-Carlo experiments to see if the most commonly used information criteria and loss functions lead practitioners to choose the right specification in a conditional volatility regime switching framework. Since many practitioners use these models, it is important to assess the performance of the selection criteria, in order to guide professionals in their model selection process. This section gives details of the different experiments.

#### 3.1. Common Design: Starting Values and Numerical Method

First, we present all the features of our overall experimental framework. In all the the experiments, data are first generated following a specific DGP described in great detail with more details below. Then,



all the different models are estimated<sup>6</sup>. Finally, the selection criteria and loss functions presented in Section 2 are computed. According to the criteria, we report the percentage of selection for each specific model. Since we simulate our data, the real value of the volatility is known, but in order to come as close as possible to reality, squared errors are also used to compute the loss functions. DGPs are restricted to the LST-GARCH(1,1) and the MS(2)-GARCH(1,1) models. We focus on these models because there is some evidence in the literature that their estimation can be problematic: for example, the latent regime state estimation of MS-GARCH models. Our objective is therefore to show that these problems can lead to a deterministic regime switching type being chosen instead of a stochastic one, resulting in spurious interpretations. To interpret our results, we propose a working hypothesis that defines whether a selection criterion is efficient in this simulation framework:

**Remark 1.** *From now on, we consider a selection criterion as strongly efficient if and only if it leads to the selection of the true DGP in at least 90% of cases. Moreover, we consider a selection criterion as weakly efficient if and only if it leads to selection of the true DGP regardless of the assumption on the distribution of the error term in at least 90% of cases.*

The first part of our working hypothesis is very restrictive. In this case, we consider a selection criterion as efficient only if the entire true DGP is selected. The second part relaxes the misspecification on the error distribution. For example, if the true DGP is the MSG-GARCH-H and the selection criterion leads to selection of the MST-GARCH-H in 95% of cases, we conclude that it is only weakly efficient. Each simulation experiment is replicated 2000 times with a sample size<sup>7</sup>  $T = 2000$ . With the estimation framework described in the previous section, we face the well-known issue of selecting the starting values to initialize the local optimization procedure. For the likelihood function, we use traditional methods. The sample data variance is used to initialize the conditional variance and we employ ergodic probabilities to set the parameters governing the Markov Chain for MS-GARCH models, as recommended by [29]. To handle the issue of the initial parameter values, we use the true generating values when they are known. Otherwise, starting values for the parameters are selected from a grid of ten different random sets. We specify a plausible range for the parameters to avoid unusual sets. Thus, we draw the vectors  $\tilde{\theta}_g = (\omega_1, \alpha_1, \beta_1, \nu)'$ ,  $\tilde{\theta}_{lst} = (\omega_1, \alpha_1, \alpha_2, \beta_1, \gamma, \nu)'$ ,  $\tilde{\theta}_{gjr} = (\omega_1, \alpha_1, \alpha_2, \beta_1, \gamma, \nu)'$ ,  $\tilde{\theta}_{eg} = (\omega_1, \alpha_1, \alpha_2, \beta_1, \gamma, \nu)'$  and  $\tilde{\theta}_{MS} = (\omega_1, \omega_2, \alpha_{1,1}, \alpha_{1,2}, \beta_1, \beta_2, p_{11}, p_{22}, \nu)'$  which are respectively the vectors of starting values of the GARCH, LST-GARCH, GJR-GARCH, EGARCH and MS-GARCH models used in the estimation. All the parameters are drawn using a uniform distribution:  $\omega_j \sim U(0.0001; 0.5)$ ,  $\alpha_1 \sim U(0.1; 0.3)$ ,  $\alpha_2 \sim U(-0.15; 0.15)$ ,  $\alpha_{1,j} \sim U(0.1; 0.3)$ ,  $\beta_j \sim U(0.4; 0.9)$ ,  $\gamma \sim U(0; 5)$  and  $p_{jj} \sim U(0; 1)$ . Parameter  $\nu$  is null when data are estimated under the normality assumption. In the Student case,  $\nu$  also follows a uniform distribution such that  $\nu \sim U(2.1; 10)$ . For each set of starting values, we check the positivity and the stationarity of the process. If the constraints

<sup>6</sup> For each experiment, we estimate these models: GARCH, GARCH-T, LST-GARCH, LST-GARCH-T, GJR-GARCH, GJR-GARCH-T, EGARCH, AEGARCH-T, MSG(2)-GARCH-H, MSG(2)-GARCH-K, MST(2)-GARCH-H and MST(2)-GARCH-K.

<sup>7</sup> Results for  $T = 1000$  are available on demand, results remain the same. We do not consider smaller sample size since in financial application, we used to study daily data.

are not verified, we draw another set of starting values. The set which gives the highest likelihood is selected to start the estimation procedure, which is based on the `fmincon` MATLAB routine. The options remain the same for all estimations. We also check for optimization convergence by checking the output structure of the `fmincon` function. If the procedure does not converge, we try another set of starting values until optimization convergence is achieved.

### 3.2. Experiment 1: Simulation of MS-GARCH-K Processes

The purpose of the first experiment is to generate data following an MS-GARCH-K process. As mentioned in Section 2.1.3, there are two different approaches to an MS-GARCH process. This experiment focuses on the traditional approach, the one adopted in [25]. Data are generated following Equation (12). The simulation procedure is the following. We start by generating the random error vector  $\{\eta_t\}$  and the regime process vector  $\{\Delta_t\}$  for  $t = 1, \dots, T$ . The random errors are drawn randomly from a Gaussian distribution<sup>8</sup>. The regime process variable follows a Markov chain with finite state space  $S = \{1, 2\}$  and a  $2 \times 2$  transition probability matrix  $P = \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix}$ . Next, we construct<sup>9</sup>  $h_1(\Delta_0) = \omega(\Delta_0) + \alpha(\Delta_0)\epsilon_0^2 + \beta(\Delta_0)h_0$  and  $\epsilon_1 = \eta_1\sqrt{h_1}$ . Finally, we compute recursively the sequence  $\{h_2, \epsilon_2, \dots, h_T, \epsilon_T\}$  where  $h_t = \omega(\Delta_t) + \alpha(\Delta_t)\epsilon_{t-1}^2 + \beta(\Delta_t)h_{t-1}$  and  $\epsilon_t = \eta_t\sqrt{h_t}$ . In this approach  $h_T$  depends on the entire history of regimes up to  $T$ . The variance specification of Klaassen [25] is reasonable since, as in the standard GARCH model, past shocks and past variances affect today's variances. However, it is not in keeping with the initial aim of GARCH models, which is the representation of a high order ARCH process. In this first experiment, we use four different transition matrices and two sets of GARCH parameters. For the first three cases, we have  $\omega = \begin{pmatrix} 0.001 \\ 0.05 \end{pmatrix}$ ,  $\alpha_1 = \begin{pmatrix} 0.2 \\ 0.1 \end{pmatrix}$  and  $\beta = \begin{pmatrix} 0.4 \\ 0.85 \end{pmatrix}$ . We focus on three different transition matrices  $P_1 = \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix}$ ,  $P_2 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$  and  $P_3 = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$ . In this framework, the long-run (*i.e.*, unconditional) probabilities<sup>10</sup> are equal for each process. However, the regime persistence  $\delta$ <sup>11</sup> differs. Matrix  $P_1$  represents the case where regime switches occur often (anti-persistence), while the second matrix,  $P_2$ , represents the case where the probability of remaining in the same regime is equal to the probability of switching regime (short memory behavior). The last case, matrix  $P_3$  (long memory behavior) represents persistent regimes. In a fourth case, we use a different specification for both the Markov chain and the GARCH parameters. The first regime occurs very few times, with high unconditional variance. The second regime is the most common regime, with low unconditional variance :  $\omega = \begin{pmatrix} 0.1 \\ 0.05 \end{pmatrix}$ ,

<sup>8</sup> Klaassen [25] and Haas *et al.* [26], specifically address MS-GARCH models with Student-t distribution. While this case is interesting, we do not consider it in this paper, for the sake of simplicity.

<sup>9</sup> We set  $\epsilon_0 = 0$ ,  $h_0 = 1$  and  $\Delta_0 = 1$ . We generate 2000 more observations than required to minimize any starting bias.

<sup>10</sup> The probabilities of being in regime  $i = 1, 2$ . The long-run probability of the first regime:  $\pi_1$  is equal to  $\pi_1 = \frac{1-p_{22}}{2-p_{11}-p_{22}}$ .

<sup>11</sup>  $\delta$  is computed as follows:  $\delta = p_{11} + p_{22} - 1$ .

$\alpha_1 = \begin{pmatrix} 0.4 \\ 0.1 \end{pmatrix}$  and  $\beta = \begin{pmatrix} 0.9 \\ 0.4 \end{pmatrix}$  with  $P_4 = \begin{pmatrix} 0.1 & 0.1 \\ 0.9 & 0.9 \end{pmatrix}$ . The long-run probability of being in the first regime is equal to 0.1 while the long-run probability of being in the second regime is 0.9. In this setup the first regime is highly non-stationary, and can be seen as a jumps regime: at a time  $t$  a jump occurs (regime 1), but since this regime is not persistent, we go back rapidly to the normal regime (regime 2). However, the overall process is in good keeping with conditions described in [26]. We focus mainly on transition matrices because they are the source of stochastic regime switches. The probability of switching tends to one when switching rates increase. In this case, the selection criteria could fail to identify the true nature of the model.

### 3.3. Experiment 2: Simulation of MS-GARCH-H Processes

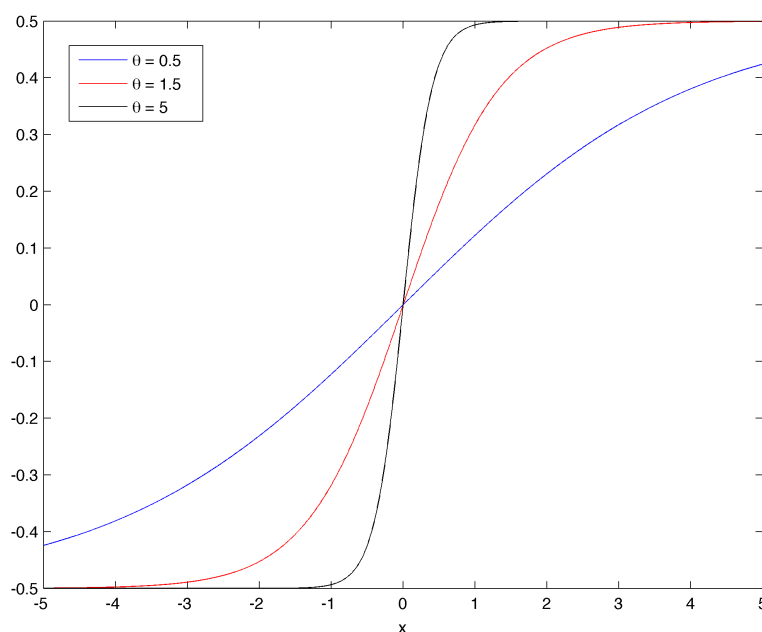
The second approach is introduced by [26] and defined by Equation (14). The simulation procedure differs from the first experiment in its second and the third steps. As before, we start by generating the random error vector  $\{\eta_t\}$  and the regime process vector  $\{\Delta_t\}$  for  $t = 1, \dots, T$ . The random errors are drawn randomly from both a Gaussian distribution and a Student distribution with a degree of freedom  $\nu = 5$ . The state variable  $\Delta_t$  follows a Markov chain with finite state space  $S = \{1, 2\}$  and a  $2 \times 2$  transition probability matrix  $P = \begin{pmatrix} p_{11} & p_{21} \\ p_{12} & p_{22} \end{pmatrix}$ . In the second step, we construct  $\mathbf{h}_1 = \omega + \alpha\epsilon_0^2 + \beta\mathbf{h}_0$ ,  $h_1(\Delta_1)$  and  $\epsilon_1 = \eta_1\sqrt{h_1(\Delta_1)}$ <sup>12</sup>. In the third step, we compute recursively the sequence  $\{h_2(\Delta_2), \epsilon_2(\Delta_2), \dots, h_T(\Delta_T), \epsilon_T(\Delta_T)\}$  where  $h_t(\Delta_t) = \omega(\Delta_t) + \alpha(\Delta_t)\epsilon_{t-1}^2 + \beta(\Delta_t)h_{t-1}(\Delta_t)$ . All specific conditional variances need to be computed at each period:  $\mathbf{h}_t = \omega + \alpha\epsilon_{t-1}^2 + \beta\mathbf{h}_{t-1}$ . Note that if the variable  $\Delta_t$  switches, the conditional variance differs instantaneously. Then, the error term is generated:  $\epsilon_t = \eta_t\sqrt{h_t(\Delta_t)}$ . The set of parameters are the same as in the first experiment, so that results can be compared. However  $\beta = \begin{pmatrix} 0.4 & 0 \\ 0 & 0.85 \end{pmatrix}$  for  $P_1$ ,  $P_2$  and  $P_3$  and  $\beta = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.4 \end{pmatrix}$  for  $P_4$  due to notation.

### 3.4. Experiment 3: Simulation of LST-GARCH Processes

In this third experiment, the underlying process is an LST-GARCH. The DGP of such a model is more trivial since there is no latent process. It is described by Equation (6) for  $p = q = 1$ . To simulate this process, we start by drawing  $\eta_t$  from a Gaussian distribution. Then, we compute recursively  $h_t$  and  $\epsilon_t$ <sup>13</sup>. As in the first experiment, we estimate the data with all the models. We focus on three different sets of parameters.  $\gamma$  is the parameter of interest since it governs the transition function. If the transition parameter is large, the transition function becomes steep and we can see it as a “two regimes” model, as shown by Figure 1. The sets of parameters are  $\omega = 0.05$ ,  $\alpha_1 = 0.3$ ,  $\alpha_2 = 0.55$ ,  $\beta = 0.3$  and  $\gamma = \{0.5, 1.5, 5\}$ .

<sup>12</sup> We set  $\epsilon_0 = 0$ ,  $h(\Delta_0)_0 = 1$  and  $\Delta_0 = 1$ . We generate 2000 more observations than required, to minimize any starting bias.

<sup>13</sup> We generate 2000 more observations than required to minimize any starting bias.



**Figure 1.** Logistic functions with different  $\gamma$ .

## 4. Results and Discussion

### 4.1. Results

Table 1 sums up the results. It shows whether or not the selection criteria are weakly efficient as defined in remark 1. Tables 2 to 12 present in detail the results of the three Monte-Carlo experiments described in Section 3. The choice percentage is given per selection criterion. For example, in Table 2, the GARCH model is selected in first position by AIC in 16.6% of cases. Our results show that when regimes are difficult to identify (*i.e.*, many regime switches or one regime occurring very few times) and data are generated as per Klaassen, information criteria and loss functions do not find always the true DGP. Thus, practitioners could make the wrong choices in these cases. Moreover, squared residuals should not be used as a proxy of volatility ; we encourage practitioners to use different proxies, like realized volatility. In addition, the results show that information criteria perform poorly when LST-GARCH is the generating process.

#### 4.1.1. Experiment 1

In this first experiment, the DGPs are MS-GARCH-K path-dependent with Gaussian innovations. With the first set of parameters, the transition probabilities are high: there are a lot of regime switches. The results of the selection process are presented in Table 1 row 1 and in Table 2. In this case, none of the criteria are strongly or weakly efficient as defined by our working hypothesis. However, if we sum the results for both MSG-GARCH-K and MST-GARCH-K, loss functions select the right model at a rate of 71.3% for MSE, 86.1% for QLIKE and 83.7% for MAE. They come close to the weakly efficient point. Loss functions computed with  $\epsilon_t^2$  often select the wrong model whatever the process used

to simulate the data, and thus are not efficient with this volatility proxy. What stands out is the poor performance of information criteria. AIC does not enable us to choose between MSG-GARCH-K and the univariate GARCH-T model, while BIC is weakly inefficient, leading to choice of a GARCH model in 98.9% of cases.

**Table 1.** Summary of the results of experiments 1 and 2. A cross means that the criterion is weakly efficient as defined in remark 1.

DGP		MSE( $h_t$ )	QLIKE( $h_t$ )	MAE( $h_t$ )	MSE( $\epsilon_t^2$ )	QLIKE( $\epsilon_t^2$ )	MAE( $\epsilon_t^2$ )	AIC	BIC
<b>Experiment 1 and 2 with Gaussian innovations</b>									
$P_1$	MS-K								
	MS-H	x	x	x				x	x
$P_2$	MS-K								
	MS-H	x	x	x				x	x
$P_3$	MS-K	x	x	x	x	x	x	x	x
	MS-H	x	x	x				x	x
$P_4$	MS-K								
	MS-H	x	x	x	x	x	x	x	x
<b>Experiment 2</b>									
$\gamma = 0.5$	LST-G								
$\gamma = 1.5$	LST-G								
$\gamma = 5$	LST-G								

**Table 2.** Results of experiment 1 with MSG-GARCH-K DGP when many switches occur. Data are generated with the transition matrix  $P_1$ . The top row of the table gives the selection criteria. The left column gives the different models. The line in grey indicates the process which should be selected in first position by each criterion. Regime persistence and volatility persistence are respectively  $\delta = -0.8$  and  $\lambda = 0.9228$ .

	MSE( $h_t$ )	QLIKE( $h_t$ )	MAE( $h_t$ )	MSE( $\epsilon_t^2$ )	QLIKE( $\epsilon_t^2$ )	MAE( $\epsilon_t^2$ )	AIC	BIC
GARCH	7.2	3.8	0.5	0	0	0	16.6	67.8
GARCH-T	4.2	2.1	0.5	0	0	0	32.2	31.1
MSG-GARCH-H	8.0	4.9	13.2	33.5	52.2	59.2	6.6	0
MST-GARCH-H	0.4	0.6	0.6	9	5.3	2.3	0.3	0
MSG-GARCH-K	30.7	46.1	54.3	49.1	34.2	34.0	27.1	0
MST-GARCH-K	40.6	40.6	29.4	8.4	8.3	4.5	1.1	0
LST-GARCH	1.1	0.9	0.1	0	0	0	0.4	0
LST-GARCH-T	1.3	0.2	0	0	0	0	1.3	0
GJR	0.7	0.4	0.1	0	0	0	3.8	0.8
GJR-T	0.7	0.3	0.1	0	0	0	5.5	0.1
EGARCH	3.1	0.1	0.4	0	0	0	2.0	0.1
EGARCH-T	2.0	0	0.8	0	0	0	2.1	0

With the matrix  $P_2$  (Table 1 row 3, Table 3), there are fewer regime switches. AIC and BIC lead to do the wrong choices at a respective rate of 57.9% and 94.6%. This frequency is a little bit lower than where there are many regime switches. The second model chosen by AIC is the MSG-GARCH-H. AIC

leads to the right choice of regime switching behavior in 82.5% of cases. The rate of regime switching errors decreases, since we less often select the GARCH model. In contrast, BIC still selects the GARCH Student specification at a rate of 82.3%. As in the previous case, the in-sample forecasts selection method improves the frequency of good choices. However, these selection criteria need to be more than weakly efficient. Loss functions computed with  $\epsilon_t^2$  return good results when data are simulated with MS-GARCH-K. The results are better than in the previous experiment: the percentage of good selection increases for all loss functions. For example, MSE leads to 49.1% good selection with the transition matrix  $P_1$ , increasing to 74.6% with  $P_2$ .

**Table 3.** Results of experiment 1 with MS-GARCH-K DGP when probabilities of remaining are equal to probabilities of switching. Data are generated with the transition matrix  $P_2$ . The top row of the table gives the selection criteria. The left column gives the different models. The line in grey indicates the process which should be selected in first position, by each criterion. Regime persistence and volatility persistence are respectively  $\delta = 0$  and  $\lambda = 0.9165$ .

	MSE( $h_t$ )	QLIKE( $h_t$ )	MAE( $h_t$ )	MSE( $\epsilon_t^2$ )	QLIKE( $\epsilon_t^2$ )	MAE( $\epsilon_t^2$ )	AIC	BIC
GARCH	0.5	0.1	0	0	0	0	0	1.8
GARCH-T	0.2	0	0.1	0	0	0	8.2	82.3
MSG-GARCH-H	24.0	17.3	23.8	13.3	21.6	19.3	21.4	5.2
MST-GARCH-H	9.3	7.0	5.4	5.6	1.7	1.5	8.4	0.2
MSG-GARCH-K	36.1	46.1	49.5	74.6	72.7	74.4	42.1	5.4
MST-GARCH-K	21.9	29.4	19.6	6.5	4.0	4.8	12.7	0.1
LST-GARCH	0	0.1	0	0	0	0	0	0
LST-GARCH-T	0.4	0	0	0	0	0	0.2	0
GJR	0.2	0	0	0	0	0	0	0.1
GJR-T	0	0	0	0	0	0	3.3	0.5
EGARCH	6.1	0	1.6	0	0	0	0	0.1
EGARCH-T	1.3	0	0	0	0	0	3.7	4.3

This improvement is accentuated when data are generated with persistent regimes. Let us consider MSG-GARCH-K and MST-GARCH-K together. In this case, all the decision criteria are weakly efficient (Table 1 row 5, Table 4). Specification errors is reduced to below 10%. All the selection criteria are efficient in this case. However, if these two models are compared, the Student version may be selected even though the data are generated with Gaussian errors.

In the fourth case, the two variances are totally different. Selection criteria lead to the choice of a misspecified model in this case (Row 7, Table 1 and Table 5). For example, BIC selects the simple GARCH model at a rate of 69.4% and the GARCH-T with at rate of 28.8%. The MSG-GARCH-K is never selected by this criterion. Loss functions do not work any better. Although MSE selects, in first position, the right model, the rate is very low (22.2%). No model is clearly selected in first position. Under the definition in remark 1, they are all inefficient and fail to recognize the true regime switching behavior.

**Table 4.** Results of experiment 1 with MSG-GARCH-K DGP when regime-specific variances are persistent. Data are generated with the transition matrix  $P_3$ . The top row of the table gives the selection criteria. The left column gives the different models. The line in grey indicates the process which should be selected in first position, by each criterion. Regime persistence and volatility persistence are respectively  $\delta = 0.8$  and  $\lambda = 0.9147$ .

	MSE( $h_t$ )	QLIKE( $h_t$ )	MAE( $h_t$ )	MSE( $\epsilon_t^2$ )	QLIKE( $\epsilon_t^2$ )	MAE( $\epsilon_t^2$ )	AIC	BIC
GARCH	0	0	0	0	0	0	0	0
GARCH-T	0	0	0	0	0	0	0	0
MSG-GARCH-H	5.1	5.5	4.4	1.9		0.1	6.4	7.9
MST-GARCH-H	3.2	4.3	1.3	0.1	0.1	0	1.7	0.2
MSG-GARCH-K	48.8	43.1	64.7	85.7	93.6	90.1	72.2	88.8
MST-GARCH-K	42.9	47.1	29.6	12.2	5.9	9.8	19.7	3.1
LST-GARCH	0	0	0	0	0	0	0	0
LST-GARCH-T	0	0	0	0	0	0	0	0
GJR	0	0	0	0	0	0	0	0
GJR-T	0	0	0	0	0	0	0	0
EGARCH	0	0	0	0.1	0	0	0	0
EGARCH-T	0	0	0	0	0	0	0	0

**Table 5.** Results of experiment 1 with MS-GARCH-K DGP when the first regime has a high variance which occurs very few times. Data are generated with the transition matrix  $P_4$ . The top row of the table gives the selection criteria. The left column gives the different models. The line in grey indicates the process which should be selected in first position, by each criterion. Regime persistence and volatility persistence are respectively  $\delta = 0$  and  $\lambda = 0.8888$ .

	MSE( $h_t$ )	QLIKE( $h_t$ )	MAE( $h_t$ )	MSE( $\epsilon_t^2$ )	QLIKE( $\epsilon_t^2$ )	MAE( $\epsilon_t^2$ )	AIC	BIC
GARCH	8.4	5.7	5.7	0	0	0	21.9	69.4
GARCH-T	7.3	5.5	5.9	0.1	0	0.1	43.8	28.8
MSG-GARCH-H	13.8	15.5	14.5	34.7	25.8	27.2	5.3	0
MST-GARCH-H	10.0	10.3	6.0	27.6	16.8	20.2	0.2	0
MSG-GARCH-K	22.2	34.2	41.7	30.0	51.9	45.3	5.0	0
MST-GARCH-K	14.9	19.7	16.7	7.4	5.2	7.0	0.5	0
LST-GARCH	3.8	1.7	1.2	0	0	0	0.2	0
LST-GARCH-T	4.8	3.3	1.8	0.2	0.2	0	1.0	0
GJR	2.3	1.1	1.5	0	0.1	0	6.9	1.4
GJR-T	2.2	2.3	2.1	0	0	0	9.9	0.1
EGARCH	5.1	0.5	1.1	0	0	0	1.9	0.2
EGARCH-T	5.2	0.2	1.8	0	0	0.1	3.4	0.1

#### 4.1.2. Experiment 2

In this second experiment, the DGP is now an MS-GARCH as per of Haas *et al.* There is a significant contrast with the previous experiment. In the first case (row 2 of Table 1 and Table 6), information criteria are all strongly efficient as defined by remark 1. As highlighted by Table 6, AIC and BIC select the MSG-GARCH-H model at a rate of 93.6% and 99.4% respectively. Loss functions computed with the true volatility  $h_t$  return a selection rate of over 50%. The MSG-GARCH-H is always selected in first



position. If the percentage selection of the MSG-GARCH-H and the MST-GARCH-H are summed, loss functions are weakly efficient. The MST-GARCH-H model provides a very good fit with data generated by an MSG-GARCH-H model.

**Table 6.** Results of experiment 2 with MS-GARCH-H DGP when many switches occur. Data are generated with the transition matrix  $P_1$ . The top row of the table gives the selection criteria. The left column gives the different models. The line in grey indicates the process which should be selected in first position, by each criterion. Regime persistence and volatility persistence are respectively  $\delta = -0.8$  and  $\lambda = 0.9228$ .

	MSE( $h_t$ )	QLIKE( $h_t$ )	MAE( $h_t$ )	MSE( $\epsilon_t^2$ )	QLIKE( $\epsilon_t^2$ )	MAE( $\epsilon_t^2$ )	AIC	BIC
GARCH	0	0	0	0	0	0	0	0
GARCH-T	0	0	0	0	0	0	0	0
MSG-GARCH-H	52.3	46.9	67.8	12.2	49.8	7.8	93.6	99.4
MST-GARCH-H	45.7	45.9	30.2	7.6	7.6	2.3	6.3	5.0
MSG-GARCH-K	0.9	3.9	1.1	65.8	30.6	70.8	0.1	0.1
MST-GARCH-K	1	3.3	0.9	12.0	12.0	19.1	0	0
LST-GARCH	0	0	0	0	0	0	0	0
LST-GARCH-T	0	0	0	0	0	0	0	0
GJR	0	0	0	0	0	0	0	0
GJR-T	0	0	0	0	0	0	0	0
EGARCH	0.1	0	0	0	0	0	0	0
EGARCH-T	0	0	0	0	0	0	0	0

With matrix  $P_2$  (Table 1 row 4, Table 7), there are again no selection problems. The rate of good selection improves in all cases. As previously, information criteria and loss functions computed with the true volatility are largely weakly efficient. However, poor performance is observed with the loss functions computed with the squared residuals. In these cases, there are errors in choosing between the two MS-GARCH-type processes.

When the regimes are persistent, the results are similar to those of the previous experiment where the DGP was the MSG-GARCH-K (Table 1 row 6, Table 8). All the criteria are efficient except when computed with volatility proxy, which leads to the selection of the MSG-GARCH-K model.

Finally, in the last special case (Table 1 row 8, Table 9), the results are very interesting; the MSG-GARCH-H is selected with a strong majority by loss functions. If we add the number of selections of the MST-GARCH-H, the selection criteria based on in-sample forecasts are weakly efficient. Information criteria seem to select the true DGP. AIC is near the strongly efficient point (89.9 % of correct selection). BIC selects the MSG-GARCH-H at a rate of 74.9%. However, it also selects the GARCH-T model in 21.8 % of replications. The poor performance of squared residuals as a proxy of volatility is again evident.

**Table 7.** Results of experiment 2 with MS-GARCH-H DGP when probabilities of remaining are equal to probabilities of switching. Data are generated with the transition matrix  $P_2$ . The top row of the table gives the selection criteria. The left column gives the different models. The line in grey indicates the process which should be selected in first position, by each criterion. Regime persistence and volatility persistence are respectively  $\delta = 0$  and  $\lambda = 0.9165$ .

	$MSE(h_t)$	$QLIKE(h_t)$	$MAE(h_t)$	$MSE(\epsilon_t^2)$	$QLIKE(\epsilon_t^2)$	$MAE(\epsilon_t^2)$	AIC	BIC
GARCH	0	0	0	0	0	0	0	0
GARCH-T	0	0	0	0	0	0	0	0
MSG-GARCH-H	57.9	54.6	76.8	34.2	57.2	39.0	94.9	99.2
MST-GARCH-H	41.1	39.5	19.1	5.1	3.2	2.8	5.0	0.6
MSG-GARCH-K	0.7	3.2	3.7	53.4	36.8	54.1	0.1	0.2
MST-GARCH-K	0.2	2.7	0.4	7.3	3.3	4.1	0	0
LST-GARCH	0	0	0	0	0	0	0	0
LST-GARCH-T	0	0	0	0	0	0	0	0
GJR	0	0	0	0	0	0	0	0
GJR-T	0	0	0	0	0	0	0	0
EGARCH	0	0	0	0	0	0	0	0
EGARCH-T	0	0	0	0	0	0	0	0

**Table 8.** Results of experiment 2 with MSG-GARCH-H DGP when regime specific variances are persistent. Data are generated with the transition matrix  $P_3$ . The top row of the table gives the selection criteria. The left column gives the different models. The line in grey indicates the process which should be selected in first position, by each criterion. Regime persistence and volatility persistence are respectively  $\delta = 0.8$  and  $\lambda = 0.9147$ .

	$MSE(h_t)$	$QLIKE(h_t)$	$MAE(h_t)$	$MSE(\epsilon_t^2)$	$QLIKE(\epsilon_t^2)$	$MAE(\epsilon_t^2)$	AIC	BIC
GARCH	0	0	0	0	0	0	0	0
GARCH-T	0	0	0	0	0	0	0	0
MSG-GARCH-H	59.9	64.6	79.2	48.7	19.1	35.9	97.9	99.8
MST-GARCH-H	39.5	32.8	19.1	13.1	2.8	7.5	2.0	0
MSG-GARCH-K	0.6	1.8	1.7	31.2	73.0	50.0	0.1	0.2
MST-GARCH-K	0	0.1	0	6.9	5.1	6.6	0	0
LST-GARCH	0	0	0	0.1	0	0	0	0
LST-GARCH-T	0	0	0	0	0	0	0	0
GJR	0	0	0	0	0	0	0	0
GJR-T	0	0	0	0	0	0	0	0
EGARCH	0	0	0	0	0	0	0	0
EGARCH-T	0	0	0	0	0	0	0	0

**Table 9.** Results of experiment 2 with MSG-GARCH-H DGP when the first regime has a high variance which occurs very few times. Data are generated with the transition matrix  $P_4$ . The top row of the table gives the selection criteria. The left column gives the different models. The line in grey indicates the process which should be selected in first position by each criterion. Regime persistence and volatility persistence are respectively  $\delta = 0$  and  $\lambda = 0.8888$ .

	MSE( $h_t$ )	QLIKE( $h_t$ )	MAE( $h_t$ )	MSE( $\epsilon_t^2$ )	QLIKE( $\epsilon_t^2$ )	MAE( $\epsilon_t^2$ )	AIC	BIC
GARCH	0	0	0	0	0	0	0	0
GARCH-T	0	0	0	0	0	0	1.3	21.8
MSG-GARCH-H	88.2	82.0	87.2	73.0	83.6	64.2	89.9	74.9
MST-GARCH-H	8.8	8.7	5.3	13.5	3.8	5.9	1.8	0.1
MSG-GARCH-K	3.0	9.3	7.5	13.2	12.5	29.7	6.3	3.0
MST-GARCH-K	0.1	0	0	0.2	0.1	0.2	0	0
LST-GARCH	0	0	0	0	0	0	0	0
LST-GARCH-T	0	0	0	0	0	0	0	0
GJR	0	0	0	0	0	0	0	0
GJR-T	0	0	0	0	0	0	0.6	0.1
EGARCH	0	0	0	0.1	0	0	0	0
EGARCH-T	0	0	0	0	0	0	0.1	0.1

#### 4.1.3. Experiment 3

We report the results of this experiment in the last three rows of Table 1, Tables 10, 11, and 12. None of the selection criteria are efficient when  $\gamma$  is low. However, in this case, loss functions computed with the true volatility yield the best results: roughly 40% correct selection for MSE, QLIKE and MAE loss functions, regardless of the assumption on distribution. The volatility proxies are clearly inefficient and, importantly, they lead to selection of an MS-GARCH-type process in about 80% of cases. Information Criteria also lead the wrong choice of model. BIC is strongly inefficient since it selects a GARCH(1,1) model at a rate of 96.3%. The non-linear effect affecting conditional volatility is not recognized in this case.

When  $\gamma = 1.5$ , the logistic function is less smooth (red line in Figure 1). Results are presented in Table 11. Loss functions computed with the true volatility have a good selection rate roughly 80%. Thus, they are close to the weakly efficient point. However, they still perform poorly when we use  $\epsilon_t^2$ . The information criteria again fail to select the true DGP, selecting the LST-GARCH model in only 23.4% of cases for AIC and 1.3% for BIC. All the other univariate GARCH-type models are selected in preference to the LST-GARCH, especially the GJR.

The number of GJR selections with both AIC and BIC is shown to increase when  $\gamma$  increases (Table 12). There is also a large non-negligible number of selections of the EGARCH model. Moreover, the performance of MSE and MAE loss functions computed with  $h_t$  tends to be worse when  $\gamma = 5$ . Despite this, errors are not made on the asymmetric effect presents in the conditional volatility. The selection criteria never select Markov switching models, except for the volatility proxy loss functions. This experiment shows that when  $\gamma$  is not large enough, the asymmetric behavior of the conditional volatility may not be detected. Moreover, when  $\gamma$  is too large, the GJR model is a good choice because it avoids the estimation issues encountered with the LST-GARCH.

**Table 10.** Selection results of experiment 3 with a smooth logistic function. Data are generated with a LST-GARCH model with  $\gamma = 0.5$ . The top row of the table gives the selection criteria. The left column gives the different models. The line in grey indicates the process which should be selected in first position, by each criterion. The volatility persistence is  $\lambda = 0.825$ .

	$MSE(h_t)$	$QLIKE(h_t)$	$MAE(h_t)$	$MSE(\epsilon_t^2)$	$QLIKE(\epsilon_t^2)$	$MAE(\epsilon_t^2)$	AIC	BIC
GARCH	4.3	9.9	6.9	0.1	0	0	56.2	96.3
GARCH-T	4.0	8.8	6.4	0.3	0	0.2	2.6	0.2
MSG-GARCH-H	1.8	2.2	4.0	29.6	21.9	13.7	1.3	0
MST-GARCH-H	2.8	2.5	5.2	1.5	3.5	25.0	0	0
MSG-GARCH-K	1.1	1.9	2.3	41.8	61.3	45.2	4	0
MST-GARCH-K	4.3	7.5	6.5	3.9	3.4	5.1	0	0
LST-GARCH	30.9	32.3	27.3	7.3	2.5	1.9	3.5	0
LST-GARCH-T	16.7	11.1	18.8	6.0	4.6	1.9	0.1	0
GJR	14.5	10.0	9.9	1.4	1.9	0.9	26.5	3.3
GJR-T	14.8	13.5	11.4	1.1	0	0.3	1.0	0
EGARCH	2.2	0.2	0.8	3.9	0.8	3.0	8.0	0.2
EGARCH-T	2.6	0.1	0.5	3.1	0.1	2.8	0.4	0

**Table 11.** Selection results of experiment 3 with a smooth logistic function. Data are generated with a LST-GARCH model with  $\gamma = 1.5$ . The top row of the table gives the selection criteria. The left column gives the different models. The line in grey indicates the process which should be selected in first position, by each criterion. Volatility persistence is  $\lambda = 0.825$ .

	$MSE(h_t)$	$QLIKE(h_t)$	$MAE(h_t)$	$MSE(\epsilon_t^2)$	$QLIKE(\epsilon_t^2)$	$MAE(\epsilon_t^2)$	AIC	BIC
GARCH	0	0	0	0.1	0	0	7.4	34.8
GARCH-T	0.1	0.4	0.1	0.2	0	0	0.6	0
MSG-GARCH-H	0	0.2	0.1	28.0	25.1	14.1	0.1	0
MST-GARCH-H	0	0.1	0.1	0.6	1.9	23.9	0	0
MSG-GARCH-K	0.1	0.2	0	31.9	58.9	41.8	0.1	0
MST-GARCH-K	0.1	0.4	0	3.2	4.0	4.2	0	0
LST-GARCH	43.4	63.5	43.1	10.4	5.3	3.2	23.4	1.3
LST-GARCH-T	33.0	19.8	34.5	8.0	2.3	1.8	2.0	0
GJR	9.7	6.0	9.2	1.5	1.2	1.2	46	41.3
GJR-T	11.1	9.3	10.8	2.7	0	0.9	3.0	0.1
EGARCH	1.4	0.7	1.3	5.9	1.2	4.7	16.3	22.5
EGARCH-T	1.1	0.4	0.8	7.5	2	4.2	1.1	0

**Table 12.** Selection results of experiment 3 with a smooth logistic function. Data are generated with a LST-GARCH model with  $\gamma = 5$ . The top row of the table gives the selection criteria. The left column gives the different models. The line in grey indicates the process which should be selected in first position, by each criterion. The volatility persistence is  $\lambda = 0.825$ .

	MSE( $h_t$ )	QLIKE( $h_t$ )	MAE( $h_t$ )	MSE( $\epsilon_t^2$ )	QLIKE( $\epsilon_t^2$ )	MAE( $\epsilon_t^2$ )	AIC	BIC
GARCH	0	0	0	0.1	0	0	0	0.1
GARCH-T	0	0	0	0	0	0	0	0
MSG-GARCH-H	0	0	0	33.4	33.2	22.8	0	0
MST-GARCH-H	0	0	0	0.8	1.0	19.7	0	0
MSG-GARCH-K	0	0	0	30.0	47.3	37.0	0	0
MST-GARCH-K	0	0	0	4.4	5.4	3.9	0	0
LST-GARCH	48.0	62.8	49.5	7.2	3.8	2.4	28.9	1.8
LST-GARCH-T	19.6	26.0	26.3	8.8	8.4	6.3	1.3	0
GJR	18.1	5.3	13.5	6.3	0.8	3.2	55.8	92.8
GJR-T	14.3	5.9	10.7	8.4	0.1	4.5	3.1	0.3
EGARCH	0	0	0	0.4	0	5.7	10.9	5.0
EGARCH-T	0	0	0	0.2	0	0.2	0	0

#### 4.2. Discussion

To summarize the results presented in Section 4.1, we point out three main selection issues. The first one is whether Student models can outperform Gaussian models for in-sample forecasts. The second is that information criteria perform poorly in some special cases. Finally, we find that squared standardized residuals are a bad proxy for volatility when making in-sample forecasts.

The first issue is interesting. Since the Student distribution tends to the Gaussian distribution when the degree of freedom tends to infinity, it is not surprising that the MST-GARCH can fit MSG-GARCH data well. Moreover, there is no problem with interpretation. Even when a misspecification error is made on the distribution, it does not impact regime switching behavior.

We propose two possible explanations for the poor results obtained with some selection criteria. First, when two processes are similar, the likelihood of model misspecification may be higher. The likelihood value is the principal component of information criteria. MS-GARCH processes with  $P_1$  are anti-persistent since  $\lambda = -0.8$  and the probabilities of switching are equal to 0.9. In asymmetric models, regime switches are deterministic and depend on a transition function. Thus, the equation for  $h_t$  changes in each period and the probability of switching is equal to one. It is not surprising that the univariate model fits well with MS-GARCH processes with a transition matrix of the form of  $P_1$ . As shown by Guegan and Rioublanc [38], MS-GARCH models with this type of persistence have a similar autocorrelation function to the GARCH process. Moreover, the loss functions work better for the simple reason that the number of parameters impact the information criteria.

A second possible explanation is that, in mixture models, when there is a very small class that is hard to identify or when two classes are very similar, estimation may be less efficient, as pointed out by [39]. This is observed from the results of Tables 2, 5, 6, 9, Figures 2 and 3. The four tables correspond to cases where the long-run probabilities are low for at least one regime. Figure 2 gives box plots for each parameter of MSG-GARCH models when the model estimate corresponds to the DGP. This figure

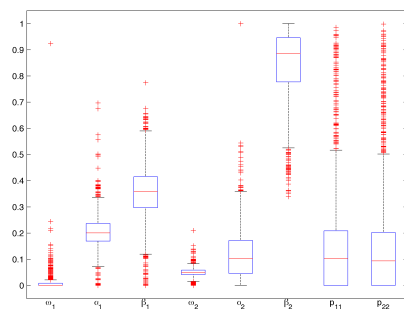
highlights the poor estimation efficiency of MSG-GARCH-K parameters when the transition matrices are  $P_1$  or  $P_4$ . Estimations are around the DGP values for both MS-GARCH models, but the variance is higher with MSG-GARCH-K (Figures 2(a) and 2(b)). Moreover, there are more extreme values in this latter process, particularly for the estimation of transition probabilities. With  $P_3$ , estimation errors are less frequent. Figure 3 represents the non-parametric density estimation<sup>14</sup> of the simulated conditional volatility processes  $\hat{f}(h_t)$  (solid lines) and the estimated conditional volatility processes  $\hat{f}(\hat{h}_t)$  (dash-dash lines), for one replication<sup>15</sup> with  $f$  the probability density function. Figure 3 shows the differences between the non parametric density functions when we use  $P_1$  and  $P_3$ . They point to a large difference between the two MSG-GARCH models, which can be explained by the construction of the MSG-GARCH-K. Although the computation of past conditional variance depends only on the previous state,  $h_t$  is dependent on all the previous regimes. In the MSG-GARCH-H, regime-specific variances are totally independent. That is why both simulated MSG-GARCH-H volatility processes exhibit bimodal densities (blue solid line). Moreover,  $\hat{f}(\hat{h}_t)$  seems to fit well  $\hat{f}(h_t)$ . With MS-GARCH-K processes,  $\hat{f}(h_t)$  functions (red solid lines) also have two modes but they are less apparent and wider. This is probably due to path-dependence behavior. Although the approximation of Klaassen is attractive, it sometimes fails to adequately fit the overall volatility time series of the path-dependent MSG-GARCH-K model; we can see that  $\hat{f}(\hat{h}_t)$  (red dash-dash line) does not exhibit the two modes when the regimes switch a lot (Figure 3(a)). Of course, these figures represent only one replication, but things are very similar for the others. All this illustrates why BIC does not choose the right model (results of Tables 2, 3 and 7): if the regimes are not properly identified, the estimation will be less accurate and the value of the likelihood function lower. Thus, AIC and BIC have higher values and become inaccurate in selecting the model. The loss functions computed with the true volatility are not affected by the likelihood value and are more efficient than the information criteria.

Finally, the impact of using  $\epsilon_t^2$  differs depending on the loss function and the DGP. For example, the frequency of good choices indicated by the QLIKE loss function decreases when regimes become more persistent and the DGP is the MSG-GARCH-H, whereas it increases when the DGP is the MSG-GARCH-K. However, this proxy works better than information criteria when data are simulated by MSG-GARCH-K. Other proxies which are unbiased and contain much more information should give better results, so we encourage practitioners to use them to estimate RS-GARCH models. Similar remarks are made in [36,37,40] regarding the selection of models using an out-sample forecasts approach. The authors cited above recommend the realized volatility for example.

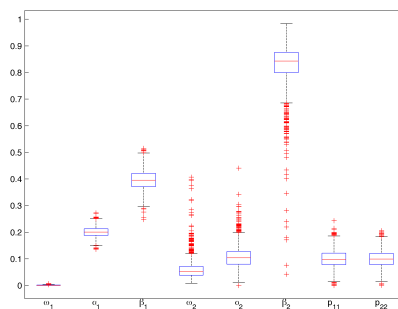
Our two experiments yield three main findings: first, when data are simulated in the sense of Haas, information criteria are a powerful means of selecting the right model. Secondly, loss functions seem to work well when the DGP is MSG-GARCH-K. Finally, although  $\epsilon_t^2$  is not a good proxy, it gives good results when data are generated in the sense of Klaassen. Our results also highlight the strong impact that estimation has on the efficiency of these criteria.

<sup>14</sup> Estimation computed with Gaussian kernel and Silverman's rule of thumb.

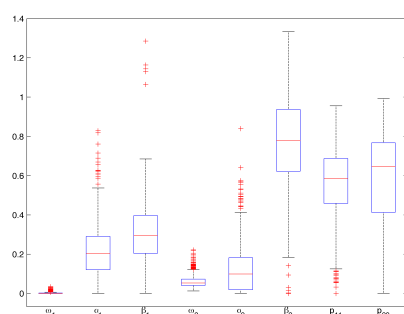
<sup>15</sup> Figure 3(a) is related to the 40th replication of the first and the second experiments with matrix  $P_1$ , BIC selects the right specification when data are simulated with MSG-GARCH-H but it selects the GARCH model for data simulated with MS-GARCHG-K. Figure 3(b) is related to the 66th replication of the first and second experiments with  $P_3$  where there is no selection problem.



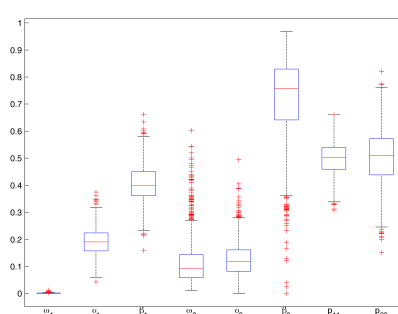
(a)



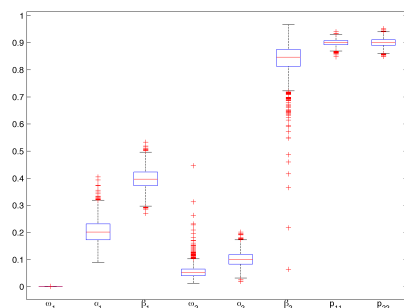
(b)



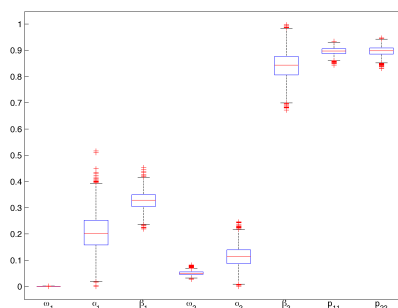
(c)



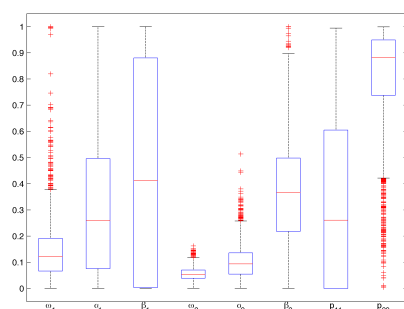
(d)



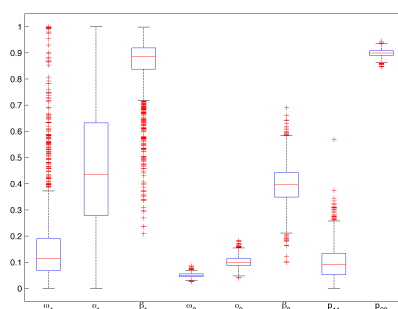
(e)



(f)



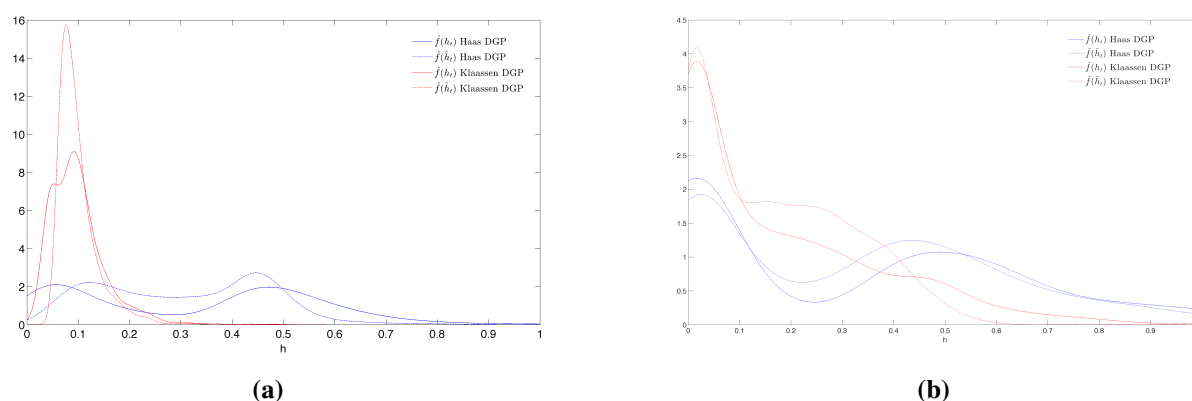
(g)



(h)

**Figure 2.** Box plots of MSG-GARCH parameter estimations. (a) MS-GARCH-K with  $P_1$ ; (b) MS-GARCH-H with  $P_1$ ; (c) MS-GARCH-K with  $P_2$ ; (d) MS-GARCH-H with  $P_2$ ; (e) MS-GARCH-K with  $P_3$ ; (f) MS-GARCH-H with  $P_3$ ; (g) MS-GARCH-K with  $P_4$ ; (h) MS-GARCH-H with  $P_4$ .

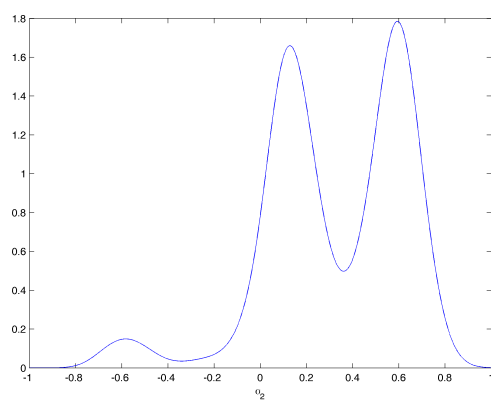
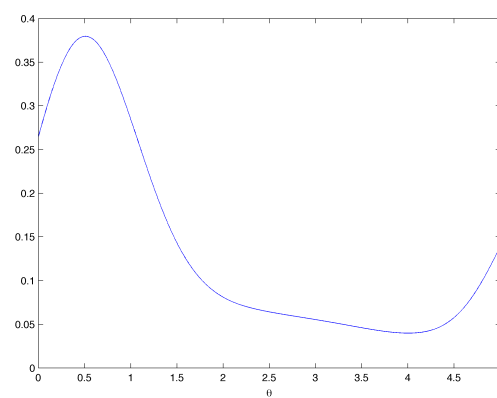
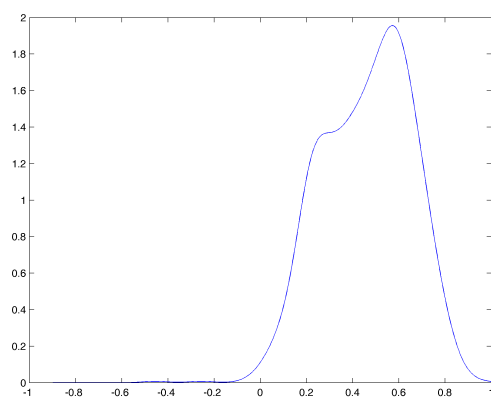
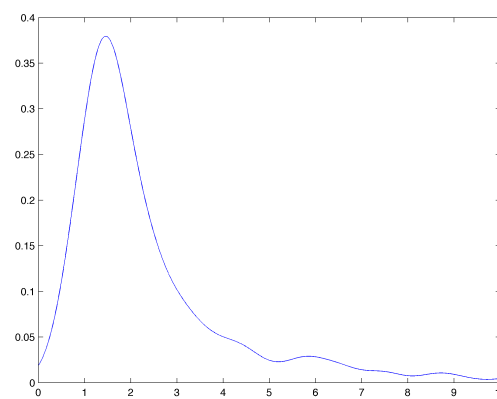
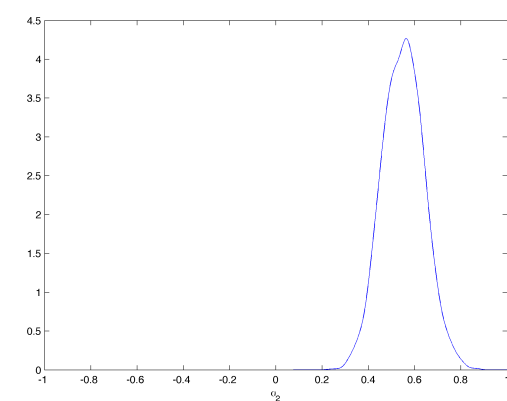
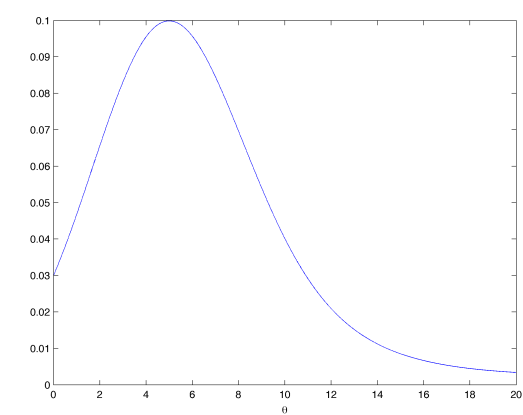




**Figure 3.** Non-parametric density estimation of the simulated and estimated conditional volatility for one replication in Experiment 1 and 2. **(a)** Transition matrix  $P_1$ ; **(b)** Transition matrix  $P_3$ .

With the third experiment, our framework reveals estimations of  $\alpha_2$  and  $\gamma$  to be very imprecise, as represented in Figure 4. This Figure gives the non-parametric density estimation<sup>16</sup> of  $\alpha_2$  and  $\gamma$  for the three cases that we investigate. Figure 4(a) shows that  $\alpha_2$  is very poorly estimated when the transition parameter is low; although there is a mode around 0.55, the true value of  $\alpha_2$ , a second mode appears around 0.1. More surprisingly, there is a third mode around  $-0.55$ , *i.e.*, we sometimes estimate an opposite asymmetric effect. A plausible explanation is that when the transition parameter is low, the logistic function is substantially flatter. As a result, estimating the coefficient attached to the logistic function and transition parameter is harder. Figures 4(c) and 4(e) show that estimation works better when  $\gamma$  increases. Figures 4(b), 4(d) and 4(f) give the non-parametric density estimations of  $\gamma$ . They highlight the well-stylized fact that QML is not an accurate means estimating of this parameter. Despite this estimation problem, regime selection is good; on average, the criteria select at least the right regime switching behavior: MS-GARCH models do not capture the continuum of regimes introduced by Gonzalez-Rivera [13]. However, the GJR model adequately fits the LST-GARCH process. As for using squared residuals as a proxy for volatility, we do not recommend this any more than in the previous experiment.

<sup>16</sup> Estimation computed with Gaussian kernel and Silverman's rule of thumb.

(a)  $\hat{\alpha}_2$ (b)  $\hat{\gamma}$ (c)  $\hat{\alpha}_2$ (d)  $\hat{\gamma}$ (e)  $\hat{\alpha}_2$ (f)  $\hat{\gamma}$ 

**Figure 4.** Non-parametric density estimation of  $\alpha_2$  and  $\gamma$  for the three sets of parameters. (a,b) correspond to the first set  $\gamma = 0.5$ ; (c,d) correspond to the second set  $\gamma = 1.5$ ; and (e,f) correspond to the third set  $\gamma = 5$ .

## 5. Conclusions

This paper presents simulation results regarding the properties of model selection criteria in a regime switching conditional volatility framework. Such models are often difficult to estimate due to their complex forms, posing a number of challenges, MS-GARCH models need to estimate many parameters. ST-GARCH models need to estimate a transition parameter, which is complicated. For example, Chan and McAleer [41] investigate the finite sample properties of MLE for Smooth Transition Autoregressive models (STAR) with GARCH component models. They show that the variability of the threshold value depends on the magnitude of unconditional shocks for the Logistic STAR model. They also examine misspecification on the transition function. Moreover, estimation by QML is known to be very sensitive to starting values.

This article contributes to three strands of the literature. First, we show that it is rare to make an error in the selection of regime switching behavior. Selection criteria manage to distinguish well between stochastic and deterministic switches in most cases. However, Information Criteria could lead practitioners to make the wrong choice between stochastic and deterministic regime switches. Loss functions computed with true volatility are more suitable for this purpose. The second contribution of this work concerns the selection of models based on selection criteria such as AIC, BIC or in-sample forecast performance. Through statistical analysis, we provide empirical evidence that results obtained using these selection criteria are directly impacted by estimation performance. Proper estimation of these models is important, and although regime switching GARCH models provide good indicators to explain the financial crisis<sup>17</sup>, selection needs to be made with great care. Estimation methods other than QML are available, like Bayesian or GMM estimations. However, with MS-GARCH models, ML is the one most commonly used in empirical work. Finally, we examine in-sample forecasts and show that the noisy proxy of squared returns is, in a lot of cases, not a good choice when selecting a volatility model.

In the first experiment, data are simulated following MSG-GARCH-K models with many transition matrices. Results show that BIC could lead to a majority selection of GARCH and LST-GARCH models when the regimes are not persistent or one of them occurs too often. In the same way, AIC selects MSG-GARCH-H. Loss functions improve model selection accuracy when computed with simulated volatility. Like many other authors, we find that squared errors are a very imprecise proxy for true volatility and these selection criteria do not lead to good results. In the third experiment, the underlying processes are LST-GARCH with different transition parameters. We show that the selection process is difficult when the transition parameter is not large enough or when it is too large. The GJR model may be a good candidate for such model estimations.

Our findings here reflect the complex nature of regime switching GARCH models, implying that practitioners need to perform carefully before using this type of model.

---

<sup>17</sup> Brunetti *et al.* [42] detect currency turmoil in southeast Asia with MS-GARCH models. Chang [43] uses Markov-Switching model to argue that macroeconomic variables have regime-dependent effects on stock return dynamics.

## Acknowledgments

I thank Anne Peguin-Feissolle and Emmanuel Flachaire for many valuable comments. Thanks also to Marjorie Sweetko for English language editing. Finally, I would like to thank Bilel Sanhaji and Gilles De Truchis who were always ready to help me. I am greatly indebted to participants for comments received at the CEF 2013 in Vancouver BC, the 13th OxMetrics User Conference in Aarhus and the 1st IAAE meeting in London. Finally, I would like to thank both reviewers for their insightful comments on the paper, as these comments led to an improvement of the work. Thanks also to Lu Liao, assistant editor of Econometrics, for the submission procedure which was a nice experience.

## Conflicts of Interest

The author declares no conflict of interest.

## References

1. Engle, R.F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica* **1982**, *50*, 987–1007.
2. Bollerslev, T. Generalized autoregressive conditional heteroskedasticity. *J. Econom.* **1986**, *31*, 307–327.
3. Lamoureux, C.G.; Lastrapes, W.D. Persistence in Variance, Structural Change, and the GARCH Model. *J. Bus. Econ. Stat.* **1990**, *8*, 225–34.
4. Box, G.E.P. Robustness in the strategy of scientific model building. In *Robustness in Statistics*; Launer, R. L., Wilkinson, G. N., Eds.; Academic Press: New York, NY, USA, 1978, pp. 201–236.
5. Gray, S.F. Modeling the conditional distribution of interest rates as a regime-switching process. *J. Financ. Econ.* **1996**, *42*, 27–62.
6. Hagerud, G.E. A Smooth Transition ARCH Model for Asset Returns. *Working Paper Series in Economics and Finance* 162; Stockholm School of Economics: Sweden, Stockholm, 1997.
7. Francq, C.; Zakoïan, J.M. Deriving the autocovariances of powers of Markov-switching GARCH models, with applications to statistical inference. *Comput. Stat. Data Anal.* **2008**, *52*, 3027–3046.
8. Bauwens, L.; Preminger, A.; Rombouts, J.V.K. Theory and inference for a Markov switching GARCH model. *Econom. J.* **2010**, *13*, 218–244.
9. Lubrano, M. Smooth Transition Garch Models: A Baysian Perspective. In *Discussion Papers (REL—Recherches Economiques de Louvain)* 2001032; Institut de Recherches Economiques et Sociales (IRES), Université Catholique de Louvain: Louvain, Belgium, 2001.
10. Chan, F.; McAleer, M.; Medeiros, M.C. Structure and asymptotic theory for nonlinear models with GARCH errors. *EconomiA* **2015**, doi:10.1016/j.econ.2015.01.001.
11. Maugeri, N. Some Pitfalls in Smooth Transition Models Estimation: A Monte Carlo Study. *Comput. Econ.* **2014**, *44*, 339–378.
12. Augustyniak, M. Maximum likelihood estimation of the Markov-switching GARCH model. *Comput. Stat. Data Anal.* **2014**, *76*, 61–75.
13. Gonzalez-Rivera, G. Smooth-Transition GARCH Models. *Stud. Nonlinear Dyn. Econom.* **1998**, *3*, 1–20.

14. Berkes, I.; Horváth, L. The efficiency of the estimators of the parameters in GARCH processes. *Ann. Statist.* **2004**, *32*, 633–655.
15. Francq, C.; Zakoian, J.M. On Efficient Inference in GARCH Processes. In *Dependence in Probability and Statistics*; Bertail, P., Soulier, P., Doukhan, P., Eds.; Springer: New York, NY, USA, 2006; Volume 187, *Lecture Notes in Statistics*, pp. 305–327.
16. Black, F. Studies of stock price volatility changes. In Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section, Boston, MA, USA, 23–26 August, 1976; pp. 177–181.
17. Ding, Z.; Granger, C.W.; Engle, R.F. A long memory property of stock market returns and a new model. *J. Empir. Financ.* **1993**, *1*, 83–106.
18. Hans, F.P.; Dijk, D.V. Forecasting stock market volatility using (nonlinear) GARCH models. *J. Forecast.* **1996**, *15*, 229–235.
19. Loudon, G.F.; Watt, W.H.; Yadav, P.K. An empirical analysis of alternative parametric ARCH models. *J. Appl. Econom.* **2000**, *15*, 117–136.
20. Glosten, L.R.; Jagannathan, R.; Runkle, D.E. On the Relation between the Expected Value and the Volatility of the Nominal Excess Return on Stocks. *J. Financ.* **1993**, *48*, 1779–1801.
21. Nelson, D.B. Conditional Heteroskedasticity in Asset Returns: A New Approach. *Econometrica* **1991**, *59*, 347–370.
22. Rockinger, M.; Jondeau, E. Entropy densities with an application to autoregressive conditional skewness and kurtosis. *J. Econom.* **2002**, *106*, 119–142.
23. Hamilton, J.D.; Susmel, R. Autoregressive conditional heteroskedasticity and changes in regime. *J. Econom.* **1994**, *64*, 307–333.
24. Cai, J. A Markov Model of Switching-Regime ARCH. *J. Bus. Econ. Stat.* **1994**, *12*, 309–316.
25. Klaassen, F. Improving GARCH volatility forecasts with regime-switching GARCH. *Empir. Econ.* **2002**, *27*, 363–394.
26. Haas, M.; Mittnik, S.; Paolella, M.S. A New Approach to Markov-Switching GARCH Models. *J. Financ. Econom.* **2004**, *2*, 493–530.
27. Gallo, G.M.; Otranto, E. Forecasting Realized Volatility with Changing Average Volatility Levels. *Int. J. Forecast.* **2015**, forthcoming.
28. Hamilton, J.D. A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle. *Econometrica* **1989**, *57*, 357–384.
29. Hamilton, J.D. Nonlinearities and the Macroeconomic Effects of Oil Prices. *Macroecon. Dyn.* **2011**, *15*, 364–378.
30. Ardia, D. Bayesian estimation of a Markov-switching threshold asymmetric GARCH model with Student-*t* innovations. *Econom. J.* **2009**, *12*, 105–126.
31. Haas, M.; Mittnik, S.; Paolella, M.S. Asymmetric multivariate normal mixture GARCH. *Comput. Stat. Data Anal.* **2009**, *53*, 2129–2154.
32. Hu, L.; Shin, Y. Optimal Test for Markov Switching GARCH Models. *Stud. Nonlinear Dyn. Econom.* **2008**, *12*, 3, doi:10.2202/1558-3708.1528.

33. Akaike, H. A New Look at the Statistical Model Identification. In *Selected Papers of Hirotugu Akaike*; Parzen, E., Tanabe, K., Kitagawa, G., Eds.; Springer Series in Statistics, Springer: New York, NY, USA, 1998; pp. 215–222.
34. Kullback, S.; Leibler, R.A. On Information and Sufficiency. *Ann. Math. Stat.* **1951**, *22*, 79–86.
35. Hurvich, C.M.; Tsai, C.L. Regression and Time Series Model Selection in Small Samples. *Biometrika* **1989**, *76*, 297–307.
36. Hansen, P.R.; Lunde, A. Consistent ranking of volatility models. *J. Econom.* **2006**, *131*, 97–121.
37. Patton, A.J. Volatility forecast comparison using imperfect volatility proxies. *J. Econom.* **2011**, *160*, 246–256.
38. Guegan, D.; Rioublanc, S. Regime Switching Models: Real or Spurious Long Memory ? Available online: <http://econpapers.repec.org/paper/haljournal/halshs-00189208.htm> (accessed on 13 January 2015).
39. Frühwirth-Schnatter, S. *Finite Mixture and Markov Switching Models*; Springer Series in Statistics, Springer: New York, NY, USA, 2006.
40. Laurent, S.; Rombouts, J.V.; Violente, F. Consistent ranking of multivariate volatility models. In *CORE Discussion Papers 2009002*; Université Catholique de Louvain, Center for Operations Research and Econometrics (CORE): Louvain, Belgium, 2009.
41. Chan, F.; McAleer, M. Maximum likelihood estimation of STAR and STAR-GARCH models: Theory and Monte Carlo evidence. *J. Appl. Econom.* **2002**, *17*, 509–534.
42. Brunetti, C.; Scotti, C.; Mariano, R.S.; Tan, A.H. Markov switching GARCH models of currency turmoil in Southeast Asia. *Emerg. Mark. Rev.* **2008**, *9*, 104–128.
43. Chang, K.L. Do macroeconomic variables have regime-dependent effects on stock return dynamics? Evidence from the Markov regime switching model. *Econ. Modell.* **2009**, *26*, 1283–1299.