Selection Criteria in Regime Switching Conditional Volatility Models

Thomas Chuffart

Selection criteria in regime switching conditional volatility models

Thomas CHUFFART*

14/07/2013

Abstract A large number of non linear conditional heteroskedastic models have been proposed in the literature and practitioners do not have always the tools to choose the correct specification. In this article, our main interest is to know if usual choice criteria lead them to choose the good specification in regime switching framework. We focus on two types of models: the Logistic Smooth Transition GARCH model and the Markov-Switching GARCH models. Thanks to simulation experiments, we highlight that information criteria and loss functions can lead practitioners to do a misspecification. Indeed, depending on the Data Generating Process used in the experiment, the choice of a criteria to select a model is a difficult issue. We argue that if selection criteria lead to choose the wrong model, it's rather due to the difficulty to estimate such models with Quasi Maximum Likelihood Estimation method (QMLE).

Keywords Conditional volatility, model selection, GARCH, Regime Switching

1 Introduction

Engle (1982) developed the Autoregressive Conditional Heteroskedasticity (ARCH) models which provide a fruitfulness framework to analyze volatility and financial time series. From this finding, it is one of the most active research topics in financial econometrics. In 1986, Bollerslev proposed Generalized Autoregressive Conditional Heteroskedasticity (GARCH) models. In GARCH modeling, the volatility is a linear function of past volatility and squared residual past shocks. These models are of the form $\varepsilon_t = \eta_t \sqrt{h_t}$ where h_t is a positive process (the volatility) and η_t an identically and independently distributed random variable with zero mean and a unit variance. However, some empirical works have shown that these models does not always adequately fit the data over a long period of time. For example, Lamoureux and Lastrapes

*Aix-Marseille University (Aix Marseille School of Economics), CNRS & EHESS
GREQAM (UMR 7316) Centre de la Vieille Charite 2, rue de la Charite 13236 Marseille cedex 02 France.
E-mail: thomas.chuffart@univ-amu.fr. Tel: +33 786 46 47 58

(1990) show that if structural changes are not considered, it may bias upward GARCH estimates of persistence in variance. To circumvent this problem, practitioners could use other types of model such as regime switching models (RS-GARCH). Several regime switching models exist: in this paper we focus on regime switches which are governed by a hidden Markov Chain and models with regimes switches governed by a transition function. The main difference between these two types of model is the nature of the regime changes: Markov Switching GARCH (MS-GARCH) models have stochastic regime switches, thereby the volatility can take different forms depending on some probabilities, while, in smooth transition models, volatility is time-varying because it depends on a transition variable through a peculiar transition function. However, when practitioners study empirical data, they have to choose a specific model to estimate and to forecast conditional volatility. A misspecification will lead to bad forecast and a not consistent estimation.

In this article, we are interested in knowing if usual choice criteria lead the practitioner to choose the good specification in regime switching framework. To provide an answer to this question, we perform Monte Carlo experiments: we simulate data following some Data Generating Processes (DGP) as LST-GARCH or MS-GARCH and then, we estimate it with different specifications. We apply selection criteria to these estimations and look at which model is chosen by the criteria. Some of our results are very surprising in the sense that, depending on the way we simulate data are simulated, selection criteria can lead to misspecification. Does that mean that these criteria are not suitable for regime switching conditional volatility models? Probably not: we argue that if these criteria lead to choose a wrong specification it is rather due to the difficulties to estimate these regime switching models with QMLE method especially when the regimes are poorly identified in MS-GARCH models.

The remainder of this article is laid out as follow. In section 2, we explain briefly the different types of models we explore. In section 3, we present our selection criteria. In section 4 we develop our simulation experiments on estimation and highlights the results. The section 5 gives some concluding remarks.

2 Models

In this section we describe the three models which are under our interest. The first one is the Logistic Smooth Transition GARCH model (LST-GARCH) model of Hagerud (1996) and González-Rivera (1998). The two other types of models belong to the class of mixture regimes GARCH models: MS-GARCH of Klaassen (2002) and of Haas et al (2004).

2.1 LST-GARCH model

This kind of model belongs to the class of "asymmetric" or "leverage" volatility models. They have been introduced by Hagerud (1996) and González-Rivera (1998) when empirical evidences have shown that the increase in volatility is larger when

returns are negative than when they are positive [1]. This characteristic is known as the "leverage-effect". The LST-GARCH model have the same dynamics as the well know GJR-GARCH model of Glosten et al (1993): the indicator discrete function is replaced by the logistic function. Particularly, this allows negative past shocks to have a bigger impact than positive past shocks on the current conditional volatility. In LST-GARCH model, the return of an asset is represented as follow for $t = 1, \ldots, T$ with T the sample size:

$$r_t = \varepsilon_t \tag{1}$$

and

$$\varepsilon_t = \eta_t \sqrt{h_t}; \tag{2}$$

ε_t is the error term, η_t an identically and independent distributed random variable with zero mean and unit variance. h_t is the conditional volatility given by,

$$h_t = \omega + \sum_{j=1}^{q} [\alpha_{1j} + \alpha_{2j} F(\varepsilon_{t-j})] \varepsilon_{t-j}^2 + \sum_{i=1}^{p} \beta_i h_{t-i} \tag{3}$$

for a LST-GARCH(p,q) process where

$$F(\varepsilon_{t-j}) = (1 + exp(\theta \varepsilon_{t-j}))^{-1} - \frac{1}{2} \tag{4}$$

with θ, the so-called transition parameter. Note that the more θ is large, the more the slope is steep. In terms of regime, because this function is continuous, González-Rivera (1998) talks about of a "continuum" of regimes where the probability to switch from one regime to another is one. However, the terms "regimes" do not have the same sense as in the two next models. Conditions to have a stationary positive process are given in Hagerud (1996) and González-Rivera (1998). This model can be estimated by maximizing the log-likelihood function given by $L = \sum_{t=1}^{T} l_t$ and

$$l_t = -\frac{1}{2} \log 2\pi - \frac{1}{2} \log h_t - \frac{1}{2} \frac{\varepsilon_t^2}{h_t} \tag{5}$$

assuming that the ε_t are normally distributed. The estimation is done with standard numerical methods but it is still dependent on the starting values used for the estimation. Moreover, the starting value of the transition parameter should be set carefully. In practice, some empirical analysis, like in Hagerud (1996), show that according to selection criteria, this model constitutes an improvement over standard GARCH models.

2.2 MS-GARCH models

The chapter three in the Handbook of Volatility Models and their Applications (Haas and Paolella (2012)) gives us a complete description of these models and their properties. These models give rise to a conditional mixture distribution and deliver an entire parametric forecast. Moreover, they allow a time-varying skewness contrary to

---

[1] See Black (1976),Ding et al (1993), Hans and Dijk (1996), Loudon et al (2000)

the traditional GARCH type models: asymmetry exists in the conditional return distribution but this asymmetry is time varying (Rockinger and Jondeau (2002)). The main difference with the previous model is the following for $t = 1, \ldots, T$:

$$r_t = \varepsilon_t \tag{6}$$

with

$$r_t = \eta_t \sqrt{h_{\Delta_t,t}} \tag{7}$$

where Δ_t is a variable which indicates the state of the world at time t. We will focus on MS-GARCH processes, that's mean Δ_t follows a Markov chain with finite state spaces $S = 1, \ldots, k$, and a transition matrix P. However, it exists other models where transition probabilities have another definition. In contrast with LST-GARCH process, probability to switch from one regime to another is no longer equal to one but depends on the transition matrix P, given by

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

with $p_{ij} = p(\Delta_t = j | \Delta_{t-1} = i)$ the probability to be in state j at time t given to be in the state i at the time $t - 1$. In this sense, this type of regime switch is endogenous. Regime specific conditional variances are

$$h_{jt} = \omega_j + \alpha_j \varepsilon_{t-1}^2 + \beta_j h_{\Delta_{t-1},t-1}, \tag{8}$$

with $j = 1, \ldots, k$

Since Hamilton and Susmel (1994) and Cai (2007) who introduced MS-ARCH processes, many models have been proposed. Gray (1996) is the first to circumvent the path dependence problem which does not allow a generalization of MS-ARCH models. If regime specific conditional variances are defined like in (8), 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. Gray introduces a 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 $h_{\Delta_{t-1},t-1}$ by the conditional variance of the error term ε_{t-1} given the information up to $t - 2$. Klaassen (2002) enlarges the information set up to $t - 1$ by conditioning the expectation of previous conditional variances on all available observations and also on the current regime:

$$h_{jt} = \omega_j + \alpha_j \varepsilon_{t-1}^2 + \beta_j \sum_{i=1}^{k} p(\Delta_{t-1} = i | \Omega_{t-1}, \Delta_t = j) h_{i,t-1}, \tag{9}$$

with $j = 1, \ldots k$ and Ω_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. He still supposes that current volatility depends on past volatility regime whatever the state of the world.

This approach contrasts with the model of Haas et al (2004) where each specific conditional variances depend only on its own lag, $j = 1, ..., k$

$$h_{jt} = \omega_j + \alpha_j \varepsilon_{t-1}^2 + \beta_j h_{j,t-1} \tag{10}$$

This model can be rewriten in matrix form:

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1}, \tag{11}$$

where $\omega = [\omega_1, \omega_2, ..., \omega_j]'$, $\alpha_1 = [\alpha_1, \alpha_2, ..., \alpha_j]'$ and $\beta = diag(\beta_1, \beta_2, ..., \beta_j)$. h_t is thereby a vector of $k \times 1$ components. Conditions are needed to guarantee the positivity of the variance like in classical GARCH framework. As shown in Haas et al (2004), equation (11) can be inverted such that $h_t = (I - \beta)^{-1} \omega + \sum_{i=1}^{\infty} \beta^{i-1} \alpha \varepsilon_{t-1}^2$ if $max\{\beta_1, \beta_2, ..., \beta_j\} < 1$. Every regimes can be represented as an $ARCH(\infty)$ which is the direct generalization of the single-regime GARCH model. This specification permits also to practitioners to make the same interpretations about the coefficients as in the single regime framework: the future variance in state j is impacted by a unit shock component $\alpha_j (1 - \beta_j)^{-1}$, the magnitude of a shock's immediately impacted on the next h_{jt} is measured by α_j, β_j being the memory in component of the variance in state j in response to this shock.

MS-GARCH model can be easily estimated by Quasi Maximum Likelihood following the work of Hamilton (1989). The log-likelihood function is given by,

$$L = \sum_{t=1}^{T} \log f(\varepsilon_t | \Omega_{t-1}) \tag{12}$$

where $f(\varepsilon_t | \Omega_{t-1})$ is the conditional density of ε_t given the process up to time t. This density is the sum of conditional regime densities weighted by the conditional regime probabilities.

Bayesian methods and Generalized Method of Moments estimation have also been developed in recent works (see Bauwens et al (2010) for example) which can circumvent the problem of local maxima. However, by testing several starting values, QML estimation is a good way to estimate this model.

3 Selection criteria and loss functions

In empirical analysis, the model selection reposes on different methods. Practitioners can use a statistical test. A specification test between LST-GARCH processes and MS-GARCH does not exist yet[2]; therefore, we will not consider this possibility in this article. A second manner to choose a specification for a model is the information criteria. We will focus on the Akaike Information Criteria (AIC) (Akaike (1974)) and the Bayesian Information Criteria (BIC) (Black (1976)). The first one is of the following form:

$$AIC = 2m - 2\log(\hat{L}) \tag{13}$$

---

[2] Hu and Shin (2008) have introduced a test procedure which test under the null the hypothesis of a GARCH process against a MS-GARCH process.

where m the number of parameters and $\hat{L}$ the maximized value of the likelihood function for the estimated model. The BIC has a similar form,

$$BIC = m\log(T) - 2\log(\hat{L}) \tag{14}$$

with m the number of parameters, T the sample size and $\hat{L}$ the maximized value of the likelihood function for the estimated model. AIC penalizes the number of parameters less strongly than does BIC. Information criteria calculate the loss of information caused by the likelihood estimation. In-sample forecasting can also be used: it consists making a comparison between the real volatility and the predicted one. Loss functions measure the difference between them. Model with the lowest loss is selected. However, the volatility is a latent variable: in practice practitioners do not observe the real volatility and they use some proxies to compute it. Since in Monte-Carlo experiment data are simulated, the real value of the volatility is known but to match the reality, the squared errors are also used to compute the loss functions. We consider three loss functions are considered: the Mean Squared Error (MSE),

$$MSE(h_t) = \frac{1}{T}\sum_{t=1}^{T}(\sigma_t - h_t)^2 \tag{15}$$

the QLIKE loss function,

$$QLIKE(h_t) = \frac{1}{T}\sum_{i=1}^{T}\left(ln(h_t) + \frac{\sigma_t}{h_t}\right) \tag{16}$$

and the Mean Absolute Error (MAE),

$$MAE(h_t) = \frac{1}{T}\sum_{t=1}^{T}|\sigma_t - h_t| \tag{17}$$

where T is the sample size, σ_t the true volatility or a proxy and h_t is the estimated volatility computed with the estimation of the parameters. The choice of these loss functions is based on Patton (2011) and his definition of a robust loss function:

Definition 1 *A loss function, L, is "robust" if the ranking of any two (possibly imperfect) volatility forecasts, h_{1t} and h_{2t}, by expected loss is the same wether the ranking is done using the true conditional variance, h_t, or some conditionally unbiased volatility proxy.*

Then, he develops a ranking of loss functions to select model in conditional volatility. MSE and QLIKE functions are robust and return the best results whereas MAE is not a robust loss function.

4 Simulation experiments

4.1 Design of the experiments

The main idea of the paper is to provide some Monte-Carlo experiments to see if the classical information criteria and loss functions lead practitioners to choose the good

specification[3]. When we have to select a conditional volatility model, it can be done in two ways. First, we can estimate model with different specifications and look at information criteria. Secondly, we can decide to choose the model using in-sample forecasts evaluation. In this part, we focus only on estimation. The Monte-Carlo experiments are the following: first data are generated following processes described in section 2. Then, the three models are estimated by QMLE method. Finally, selection criteria and loss functions presented in section 3 are computed. The model for which selection criteria are minimum is selected. The percentage of choice of a specific model done by the different criteria is reported for each experiences.

The study is limited to LST-GARCH(1,1) and MS-GARCH(1,1) with two regimes. In each experience, $T = 2000$ observations are simulated. The use of ergodic probabilities for the Markov Chain and sample data variance for the conditional variance is recommended to initialize the likelihood function. The starting values of parameters are the true generating values of these parameters when they are known. However when a misspecified model is estimated, starting values choice is more complicated: for MS-GARCH simulated processes, starting values of the LST-GARCH estimation are the following:

$$\omega = \frac{\omega_{MS,1} + \omega_{MS,2}}{2}$$

$$\alpha_1 = \frac{\alpha_{MS,1} + \alpha_{MS,1}}{2}$$

$$\alpha_2 = \frac{|\alpha_{MS,1} - \alpha_{MS,1}|}{2}$$

and

$$\beta = \frac{\beta_{MS,1} + \beta_{MS,1}}{2}$$

where $\omega, \alpha_1, \alpha_2$ and β are the parameters in the LST-GARCH(1,1) model given by equation (3). Coefficients indexed by $_{MS,j}$ are the coefficients used to simulate the data with MS-GARCH processes (Haas and Klaassen), with $j = 1, 2$ the state of the world. This choice is made on the presumption that the estimated parameters of LST-GARCH processes will be the mean of the true values of the simulated MS-GARCH. The starting value of the parameter θ is set in a region where the transition function do not take on the extreme values ; however it is well known the estimation of this parameter is really difficult. When LST-GARCH processes are simulated, we use

$$\omega_{MS,j} = \omega$$

$$\alpha_{MS,j} = \alpha_1$$

$$\beta_{MS,j} = \beta$$

---

[3] Simulations and estimations are done with MATLAB and the fmincon routine. Programs used to obtain results are available available upon request to he author. The data of the experiments are also available.

for $j = 1, 2$ the two different regimes. In contrast with previous starting values, we have the presumption that the estimated parameters of MS-GARCH on data simulated by a LST-GARCH will be very close in both regimes. In general, these starting values seem to be powerful to attempt global maxima of the likelihood function.

4.2 Experiment 1: simulations of MS-GARCH processes

The data generating processes are the MS-GARCH as follows:

$$r_t = \varepsilon_t, \tag{18}$$

and

$$\varepsilon_t = \eta_t \sqrt{h_{t,\Delta_t}} \tag{19}$$

where η_t is a random variable following a normal distribution $\mathrm{nid}(0,1)$ and

$$h_{jt} = \omega_j + \alpha_j \varepsilon_{t-1}^2 + \beta_j h_{t-1} \tag{20}$$

$j = 1, 2$ or in matrix form,

$$h_t = \omega + \alpha \varepsilon_{t-1}^2 + \beta h_{t-1} \tag{21}$$

with $\omega = \begin{pmatrix} 0.001 \\ 0.05 \end{pmatrix}$, $\alpha_1 = \begin{pmatrix} 0.2 \\ 0.1 \end{pmatrix}$, $\beta = \begin{pmatrix} 0.4 & 0 \\ 0 & 0.85 \end{pmatrix}$ when data are simulated in the sense of Haas and $\beta = \begin{pmatrix} 0.4 \\ 0.85 \end{pmatrix}$ when data are simulated with the Klaassen model. Three transition matrices are tested: $P_1 = \begin{pmatrix} 0.1 & 0.9 \\ 0.9 & 0.1 \end{pmatrix}$, $P_2 = \begin{pmatrix} 0.5 & 0.5 \\ 0.5 & 0.5 \end{pmatrix}$ and $P_3 = \begin{pmatrix} 0.9 & 0.1 \\ 0.1 & 0.9 \end{pmatrix}$. Matrix P_1 represents the case where regime switches occur often, the second one, P_2, when probabilities to stay in the same regime are equal to the probabilities to switch of regime. The last case, matrix P_3, represents persistent simulated regimes. In a fourth experiment, we use different regime specifications, one regime where the conditional variance is large but which occurs very few times and another with a lower variance but it is the most common regime:
$\omega = \begin{pmatrix} 0.1 \\ 0.05 \end{pmatrix}$, $\alpha_1 = \begin{pmatrix} 0.4 \\ 0.1 \end{pmatrix}$, $\beta = \begin{pmatrix} 0.9 & 0 \\ 0 & 0.4 \end{pmatrix}$ or $\beta = \begin{pmatrix} 0.9 \\ 0.4 \end{pmatrix}$ with $P_4 = \begin{pmatrix} 0.1 & 0.1 \\ 0.9 & 0.9 \end{pmatrix}$
The first regime occurs very few times since $p_{11} = p(\Delta_t = 1|\Delta_{t-1} = 1) = 0.1$ and $p_{21} = p(\Delta_t = 1|\Delta_{t-1} = 2) = 0.1$.

In Tables 1, 2, 3 and 4, we present the percentages of choices given by different selection criteria for these four experiments. Our results show that when regimes are difficult to identify (i.e., many regimes switches or one regime occurs few times) and data are generated in the sense of Klaassen, information criteria and loss functions do not find always the true DGP and can lead practitioners to make wrong choices.

In the first experiment, the matrix P_1 is used to simulate the data. Results in Table 1 highlight two facts. When data are simulated in the sense of Klaassen, BIC is minimum 58.45% of time for LST-GARCH, that means it is the best model to estimate

Table 1: Simulations of Markov Switching models when many switches occur.

Simulations with MS-GARCH of Haas

	LST-GARCH	**Haas**	Klaassen
MSE(h_t)	0	99.95	0.05
MSE(ε_t^2)	0	13.05	86.95
QLIKE(h_t)	0	99.90	0.10
QLIKE(ε_t^2)	0	63.55	36.45
MAE(h_t)	0	99.95	0.05
MAE(ε_t^2)	0	2.85	97.15
AIC	0	100	0
BIC	0	100	0

Simulations with MS-GARCH of Klaassen

	LST-GARCH	Haas	**Klaassen**
MSE(h_t)	5	16.60	78.40
MSE(ε_t^2)	0	50.25	49.75
QLIKE(h_t)	0.7	19.50	79.80
QLIKE(ε_t^2)	0	59.95	40.05
MAE(h_t)	0.4	14.55	85.05
MAE(ε_t^2)	0	50.05	49.95
AIC	2.5	49.35	48.15
BIC	58.45	20.35	21.20

Note: data are generated with the transition matrix P_1
Experiments are replicated 2000 times. Each loss functions
are computed with h_t the simulated volatility
and ε_t^2 the squared residuals of the simulated process.

such data. If we decide to select a model with respect to the AIC, the MS-GARCH of Haas will be chosen 49.35% of time. Loss functions compute with the true volatility seem to work better but Haas specification is still chosen about once over five. Loss functions computed with ε_t^2 select often the wrong model whatever the process used to simulate the data.

With the matrix P_2 (Table 2), there are less regime switches. Again, there are no selection problems if data are generated with MS-GARCH of Haas but if the DGP is the Klaassen MS-GARCH, information criteria still lead to do wrong choices nearly half the time. However, the frequency of wrong choices is lower than in the case with many regime switches. BIC selects now 54.65% of time (21.20% in the first case) the true specification. As in the previous case, in-sample forecasts selection method improves the frequency of good choices. In this experiment, loss functions computed with ε_t^2 return good results when data are simulated with MS-GARCH of Klaassen. It contrast with bad results obtain where the DGP is the MS-GARCH of Haas.

This improvement of good selection is accentuated when data are generated with persistent regimes: at least, all the decision criteria lead to do good choices (Table 3). The error of specification falls down at about less than 5%.

In Table 4, the results are similar to the results in Table 1. The both variances are totally different: in the first regime, the variance is very high and very low in the second. The world is not often in the high volatility regime (10%) and, as in Table 1, choice criteria lead to choose a misspecification when MS-GARCH of Klaassen is

Table 2: Simulations of MS-GARCH model when probabilities to stay are equal to probabilities to switch.

Simulations with MS-GARCH of Haas			
	LST-GARCH	**Haas**	Klaassen
MSE(h_t)	0	99.95	0.05
MSE(ε_t^2)	0	38.90	61.10
QLIKE(h_t)	0	99.95	0.05
QLIKE(ε_t^2)	0	58.50	41.50
MAE(h_t)	0	99.80	0.20
MAE(ε_t^2)	0	32.75	67.25
AIC	0	100	0
BIC	0	100	0

Simulations with MS-GARCH of Klaassen			
	LST-GARCH	Haas	**Klaassen**
MSE(h_t)	0.45	29.95	69.60
MSE(ε_t^2)	0	15.20	84.80
QLIKE(h_t)	0	26.35	73.65
QLIKE(ε_t^2)	0	24.25	75.65
MAE(h_t)	0	20.80	79.20
MAE(ε_t^2)	0	15.95	84.05
AIC	0	45.35	54.65
BIC	0	45.35	54.65

Note: data are generated with the transition matrix P_2
Experiments are replicated 2000 times. Each loss functions
are computed with h_t the simulated volatility
and ε_t^2 the squared residuals of the simulated process.

the DGP. Information criteria lead to choose a model with a different regime switches (LST-GARCH). In contrary, when the MS-GARCH model of Haas is used to generate data, every criteria work well.

In mixture model, when there is a very small class that is hard to identify or when two classes are very similar the estimation could be less efficient as noticed by Frhwirth-Schnatter (2006). This phenomena is observed with the results of Tables 1, 4 and Figures 1a, 1b. They represent the non parametric density estimation[4] of simulated and estimated volatilities for one replication[5]. They highlight a large difference between the two MS-GARCH models which can be explain by the construction of the Klaassen MS-GARCH: although the computation of the past conditional variance depends only on the previous state, h_t is dependent of all the previous regimes. In the MS-GARCH of Haas, regime specific variances are totally independent. That is why simulated and estimated volatilities of MS-GARCH model of Haas et al (2004) have bimodal densities and each modes are well distinct. The densities of the volatilities simulate and estimate with MS-GARCH of Klaassen are more surprising: when the

---

[4] Estimation computed with Gaussian kernel and Silverman's rule of thumb.

[5] Figure 1a is related to the 40th replication of the first experiment, BIC selects the good specification when data are simulated with MS-GARCH of Haas but it selects the LST-GARCH model for data simulated with MS-GARCH of Klaassen. Figure 1b is related to the 66th replication of the third experiment where there is no selection problem.

Table 3: Simulations of Markov Switching models where regime specific variances are persistent.

Simulations with MS-GARCH of Haas			
	LST-GARCH	**Haas**	Klaassen
MSE(h_t)	0	98.60	1.4
MSE(ε_t^2)	0	89.20	10.80
QLIKE(h_t)	0	94.35	5.65
QLIKE(ε_t^2)	0	20.15	79.85
MAE(h_t)	0	99.65	0.35
MAE(ε_t^2)	0	61.75	38.25
AIC	0	99.40	0.6
BIC	0	99.40	0.6

Simulations with MS-GARCH of Klaassen			
	LST-GARCH	Haas	**Klaassen**
MSE(h_t)	0	4.55	95.45
MSE(ε_t^2)	0	2.50	97.50
QLIKE(h_t)	0	5.8	94.2
QLIKE(ε_t^2)	0	0.05	99.95
MAE(h_t)	0	1.7	98.30
MAE(ε_t^2)	0	0	100
AIC	0	3.10	96.90
BIC	0	3.10	96.90

Note: data are generated with the transition matrix P_3
Experiments are replicated 2000 times. Each loss
functions are computed with h_t the simulated volatility
and ε_t^2 the squared residuals of the simulated process.

switches of regime often occur, the density is bimodal whereas when the regimes are persistent the density is unimodal but wider. In the first experiment, the modes are not really distinct (Figure 1a) and the density of the estimated volatility does not recognize these two modes. In the second one, the density of the estimated volatility seems nearest to the simulated volatility. These facts illustrate why BIC does not choose the good model (results of Tables 1 and 4): if the regimes are not well identified, the estimation will be less accurate and the value of the likelihood function lower. This leads to a higher AIC and BIC. The loss functions computed with the true volatility are not affected by the likelihood value and are more efficient than the information criteria.

Finally, depending on the loss function and the DGP, the use of ε_t^2 have a different impact. For example, the frequency of good selections done by the QLIKE loss function decreases when regimes become more persistent and the DGP is the MS-GARCH of Haas whereas it increases when the DGP is the MS-GARCH of Klaassen. However, this proxy works better than information criteria when data are simulated by Klaassen MS-GARCH. A better proxy should give better results, so we encourage practitioners to use in-sample forecasts to select RS-GARCH models in this case. Similar remarks are made in Hansen and Lunde (2006), Patton (2011) and Laurent et al (2009) in selection models using out-sample forecasts approach. The authors cited above recommend the realized volatility for example.

Table 4: Simulations of Markov Switching models when the first regime has a high variance which occurs very few times.

Simulations with MS-GARCH of Haas			
	LST-GARCH	**Haas**	Klaassen
$MSE(h_t)$	0.001	98.35	1.549
$MSE(\varepsilon_t^2)$	0	98.20	1.8
$QLIKE(h_t)$	0	99.70	0.3
$QLIKE(\varepsilon_t^2)$	0	99.8	0.2
$MAE(h_t)$	0	99.60	0.4
$MAE(\varepsilon_t^2)$	0	96.25	3.75
AIC	0	100	0
BIC	0	100	0

Simulations with MS-GARCH of Klaassen			
	LST-GARCH	Haas	**Klaassen**
$MSE(h_t)$	16.25	35.2	48.55
$MSE(\varepsilon_t^2)$	0.05	58.7	41.25
$QLIKE(h_t)$	4.45	45.30	50.25
$QLIKE(\varepsilon_t^2)$	0.05	48	51.95
$MAE(h_t)$	0.05	47.2	48.75
$MAE(\varepsilon_t^2)$	4.05	55.9	44.00
AIC	12.4	58.60	29
BIC	79.65	14.6	5.75

Note: data are generated with the transition matrix P_4
Experiments are replicated 2000 times. Each loss functions
are computed with h_t the simulated volatility
and ε_t^2 the squared residuals of the simulated process.

This experiment highlights three main things: first, when data are simulated in the sense of Haas, information criteria are powerful to select the good model among these three RS-GARCH. Secondly, loss functions seem to work well when the DGP is MS-GARCH of Klaassen. Finally, although ε_t^2 is not a good proxy, it gives good results when data are generated in the sense of Klaassen. We explain these three facts by the difficulty to estimate such models.
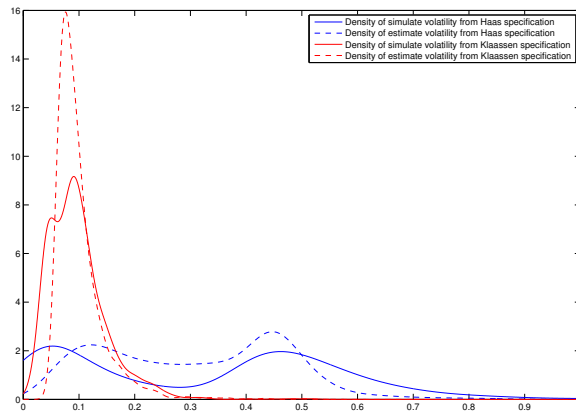
(a) Experiment 1 with transition matrix P_1
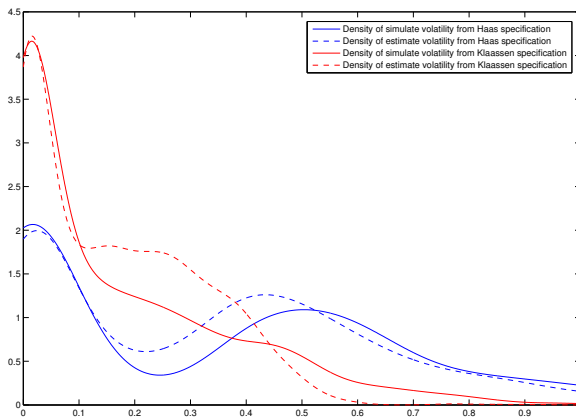
(b) Experiment 3 with transition matrix P_3

Fig. 1: Non parametric density estimation of simulate and estimate volatility for one replication in experiment 1 and 3
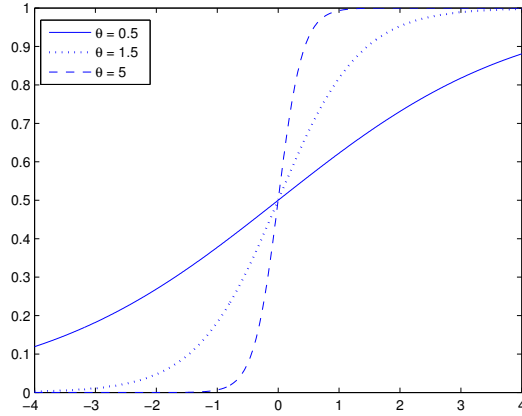
Fig. 2: Logistic functions with different θ

4.3 Experiment 2: simulations of LST-GARCH processes

In a similar way, data have been simulated following LST-GARCH processes and estimated with MS-GARCH. Data generating processes are described in section 2 by equations (1), (2) and (3). More specially,

$$h_t = \omega + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 F(\varepsilon_{t-1})\varepsilon_{t-1}^2 + \beta h_{t-1} \qquad (22)$$

with $\omega = 0.05$, $\alpha_1 = 0.3$, $\alpha_2 = 0.55$ and $\beta = 0.3$. F is the logistic function, $F(\varepsilon_{t-1}) = (1 + exp(\theta \varepsilon_{t-1}))^{-1} - \frac{1}{2}$ with $\theta = \{0.5, 5\}$. θ 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 the Figure 2.

   The results show that all information criteria and loss functions guide the practitioner to select the good specification (Table 6 and Table 5). There is though a divergence between the two experiments: a wrong model is rarely chosen if $\theta = 5$ (0% of bad selection) whereas with $\theta = 0.5$, a wrong model is chosen 16.9% of times if the AIC is used to select the model to estimate. Our experiment framework shows that the estimations of α_2 and θ seem very imprecise. This fact is represented by Figures 3 and 4. We represent on these Figures the non parametric density estimation[6] of α_2 and θ. Figure 3a shows that α_2 is very poorly estimated when the transition parameter is low: although there is a mode around 0.55, the true value of α_2, a second mode appears around 0.1. More surprising, there is a third mode around -0.55 i.e. we estimate sometimes an opposite asymmetric effect. A plausible explanation is that, when the transition parameter is low, the logistic function is substantially flat, by this fact, the estimation of the coefficient attached to the logistic function and transition parameter is harder. Figure 4 shows that the estimation is better when θ increases. Figures

---

[6] Estimation computed with Gaussian kernel and Silverman's rule of thumb.

Table 5: Simulations of LST-GARCH models with a smooth logistic function

Simulations with LST-GARCH ($\theta = 0.5$)			
	LST-GARCH	Haas	Klaassen
MSE(h_t)	90.90	4.30	4.8
MSE(ε_t^2)	12.20	44.75	43.05
QLIKE(h_t)	89.65	5.15	5.20
QLIKE(ε_t^2)	8.6	45.5	45.65
MAE(h_t)	89.35	5.40	5.25
MAE(ε_t^2)	8.6	45.45	45.95
AIC	83.1	11.5	5.4
BIC	99.75	0.25	0

Experiments are replicated 2000 times. Each loss functions
are computed with h_t the simulated volatility
and ε_t^2 the squared residuals of the simulated process.

Table 6: Simulation of LST-GARCH models with a steep logistic function

Simulations with LST-GARCH ($\theta = 5$)			
	LST-GARCH	Haas	Klaassen
MSE(h_t)	100	0	0
MSE(ε_t^2)	13.60	46.15	40.25
QLIKE(h_t)	100	0	0
QLIKE(ε_t^2)	6.25	53	40.75
MAE(h_t)	100	0	4
MAE(ε_t^2)	12.35	48.05	24.80
AIC	100	0	0
BIC	100	0	0

Experiments are replicated 2000 times. Each loss functions
are computed with h_t the simulated volatility
and ε_t^2 the squared residuals of the simulated process.

3b and 4b represent the non parametric density estimations of θ. They highlight the well stylized fact that the QMLE method of this parameter is inacurrate. Despite this estimation problem, the selection model is still good, in mean, criteria select at least about 85% of times the true model: MS-GARCH models do not capture the continuum of regimes introduced by González-Rivera (1998). Finally, results show that the loss functions computed with the proxy of the volatility do not recognize the DGP: MSE(ε_t^2) is minimum only 12.20% of times for the LST-GARCH for example with $\theta = 0.5$ according to the previous experiments when the data were generated by MS-GARCH models.
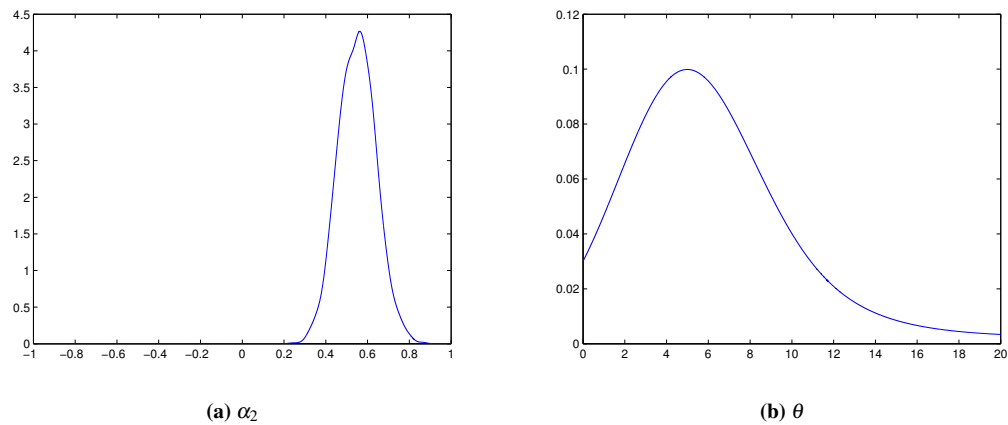
(a) α_2 **(b)** θ

Fig. 3: Non parametric density estimation of α_2 and θ when the transition function is steep

(a) α_2 **(b)** θ
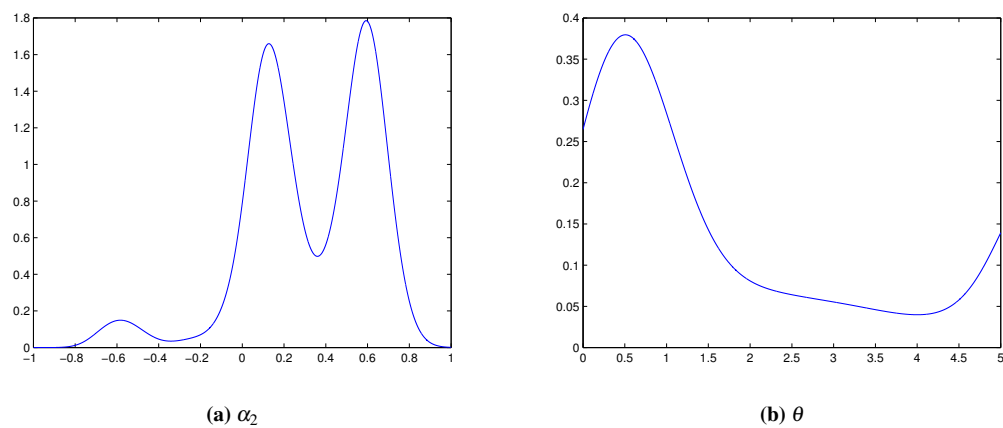
Fig. 4: Non parametric density estimation of α_2 and θ when the transition function is smooth

5 Concluding remarks

The paper presents a number of simulation results regarding the properties of selection criteria in regime switching framework in the conditional volatility. Such models are often difficult to estimate because of their complex form. MS-GARCH models need to estimate $k \times (p+q+1) + (k-1)$ parameters where k is the number of regimes and p and q are the lags of the GARCH part. Moreover, estimation by QML is very

sensitive to starting values. ST-GARCH models need to estimate a transition parameter which is a difficult issue. For example, Chan and McAleer (2002) investigate finite sample properties of MLE for Smooth Transition Autoregressive models 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 examine also the misspecification on the transition function.

Some researchers select their model based on selection criteria as AIC, BIC or in-sample forecast performance. However, results obtain by these selection criteria are directly impacted by the estimation performance. This paper examines the impact of a misspecification on selection of regime switching models in the conditional volatility. Having a good estimation of these models is an important issue. RS-GARCH models can take account structural changes and provide good indicators to explain financial crisis[7]. By Monte-Carlo experiments, we can highlight if the widely use Information Criteria are great to choose a regime switching conditional volatility model.

In the first experiment, data are simulated following MS-GARCH models of Haas and Klaassen with many transition matrices. When data are simulated in the sense of Klaassen, results show that BIC could lead to select most of time LST-GARCH model when the regimes are not persistent or one of them occurs too often. In the same way, AIC selects MS-GARCH of Haas. Loss functions improve the good selection of model when there are compute with the simulated volatility. However, since the volatility is a latent variable, researchers and practitioners can not use this measure: the use of a proxy of this volatility is needed. In this paper, we can only try the simulated squared errors. As many authors, we find that this proxy is very imprecise and selection criteria do not lead to good results.

In the second experiment, the underlying processes are LST-GARCH with different transition parameter. The paper shows that, the selection criteria lead to choose the select the right model. We note a weak improvement when this threshold is enough large.

Results present here reflect the complex nature of RS-GARCH models. Statistical analysis and statistical tests would be require before to use these models.

---

[7] Brunetti et al (2008) detect currency turmoil in southeast Asia with MS-GARCH models. Chang (2009) use Markov-Switching model to argue that macroeconomic variables have regime-dependent effects on stock return dynamics.

References

Abramson A, Cohen I (2007) On the stationarity of markov-switching garch processes. Econometric Theory 23:485–500

Akaike H (1974) A new look at the statistical model identification. IEEE Transactions on Automatic Control 19:716723

Bauwens L, Preminger A, Rombouts JVK (2010) Theory and inference for a markov switching garch model. Econometrics Journal 13(2):218–244

Black F (1976) Studies of stock price volatility changes. In: Proceedings of the 1976 Meetings of the American Statistical Association, Business and Economics Statistics Section, pp 177–181

Bollerslev T (1986) Generalized autoregressive conditional heteroskedasticity. Journal of Econometrics 31:307–327

Brunetti C, Scotti C, Mariano RS, Tan AH (2008) Markov switching garch models of currency turmoil in southeast asia. Emerging Markets Review 9(2):104–128

Cai J (2007) A markov model of switching-regime arch. Journal of Business & Economics Statistics 12(3):309–316

Chan F, McAleer M (2002) Maximum likelihood estimation of star and star-garch models: theory and monte carlo evidence. Journal of Applied Econometrics 17(5):509–534

Chang KL (2009) Do macroeconomic variables have regime-dependent effects on stock return dynamics? evidence from the markov regime switching model. Economic Modelling 26(6):1283–1299

Ding Z, Granger CWJ, Engle RF (1993) A long memory property of stock market returns and a new model. Journal of Empirical Finance 1(1):83–106

Engle RF (1982) Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. Econometrica 50(4):987–1007

Frhwirth-Schnatter S (2006) Finite mixture and Markov switching models. Springer series in statistics, Springer, New York

Glosten L, Jaganathan R, Runkle D (1993) On the relation between the expected value and the volatility of the nominal excess return on stocks. The journal of finance 48(5):1779–1801

González-Rivera G (1998) Smooth-transition GARCH models. Studies in Nonlinear Dynamics & Econometrics 3(2)

Gray SF (1996) Modeling the conditional distribution of interest rates as a regime switching process. Journal of Financial Economics 42:27–62

Haas M, Mittnik S (2006) Multivariate normal mixture GARCH. CFS Working Paper Series

Haas M, Paolella MS (2012) Mixture and regime-switching garch models. In: Handbook of volatility models and their applications

Haas M, Mittnik S, Paolella MS (2004) A New Approach to Markov-Switching GARCH Models. Journal of Financial Econometrics 2(4):493–530

Hagerud GE (1996) A smooth transition ARCH model for asset returns. Stockholm School of Economics Working Paper

Hamilton JD (1989) A new approach to the economic analysis of nonstationary time series and the business cycle. Econometrica 57(2):357–84

Hamilton JD (2011) Nonlinearities and the macroeconomic effects of oil prices. Macroeconomic Dynamics 15(S3):364–378

Hamilton JD, Susmel R (1994) Autoregressive Conditional Heteroskedasticity and changes in regime. Journal of Econometrics 64:307–333

Hans FP, Dijk Dv (1996) Forecasting stock market volatility using (nonlinear) garch models. Journal of Forecasting pp 229–235

Hansen PR, Lunde A (2006) Consistent ranking of volatility models. Journal of Econometrics (24)

Hu L, Shin Y (2008) Optimal test for markov switching garch models. Studies in Nonlinear Dynamics & Econometrics 12(3):3

Klaassen F (2002) Improving GARCH Volatility Forecasts with Regime-Switching GARCH. Empirical Economics

Lamoureux CG, Lastrapes WD (1990) Persistence in variance, structural change, and the garch model. Journal of Business & Economic Statistics 8(2):225–34

Laurent S, Rombouts J, Violante F (2009) Consistent ranking of multivariate volatility models

Loudon GF, Watt WH, Yadav PK (2000) An empirical analysis of alternative parametric arch models. Journal of Applied Econometrics 15(2):117–136

Patton AJ (2011) Volatility forecast comparison using imperfect volatility proxies. Journal of Econometrics 160(1):246–256

Rockinger M, Jondeau E (2002) Entropy densities with an application to autoregressive conditional skewness and kurtosis. Journal of Econometrics 106(1):119–142

Schwarz G (1978) Estimating the dimension of a model. The annals of statistics