Servière Adrien
El Jeilany Mohamed

ENSAE 3^{rd} year
2022-2023

Project.

Dynamic models with latent variables.

December 2022.

Contents

1 A review of the literature around Markov switching GARCH models (MS-GARCH)

1.1 Introduction on GARCH models and motivation

When it comes to modeling financial time series, researchers develop models to take stylized facts into account. One of them is that conditional volatility seems to be random. Indeed if y_t is the log-return of an asset, $(F_t)_t$ a filtration representing information up to t, then $E[y_t|F_{t-1}]$ is the best approximation of y_t in the least square sense with information up to time t. Therefore the quantity $y_t - E[y_t|F_{t-1}]$ is the error inherent to the approximation. In practice, the average of the squared error conditional to information up to t $E[(y_t - E[y_t])^2|F_{t-1}]]$ (the conditional volatility), is stochastic. This phenomenon cannot be reproduced via usual models like ARMA, because their conditional volatility is constant. Therefore, one needs to find models that make it possible to have stochastic conditional volatility. The most famous class of such models is the autoregressive conditional heteroskedasticity (ARCH) specification introduced by Engle [6] and generalized (GARCH) by Bollerslev [2]. We recall the definition of a GARCH(p,q) model

$$E[y_t|F_{t-1}] = 0 \quad \text{(martingale increment)}$$
$$V[y_t|F_{t-1}] := \sigma_t^2 = a + \sum_{1\leq i\leq p} b_i y_{t-i}^2 + \sum_{1\leq i\leq q} c_i \sigma_{t-i}^2$$

A specific way to satisfy these properties is to let $y_t = \sigma_t \eta_t$, η_t being a white noise WN(1) independent from σ.

Indeed if we want to make the volatility stochastic and dependent on previous values of the processes (y_t^2) and (σ_t^2), the simplest way is via a linear combination of the squares to make it remain positive. It also takes the volatility clustering effect into account: the higher the coefficient, the more powerful the volatility clustering. However, as seen in the article under-study, this model is no longer valid on time series with structural change.

1.2 Misspecification when calibrating a GARCH model on a sample having structural change

Indeed, even though the volatility clustering is taken into account, the conditional volatility remains too 'constant' (what is called *persistence* in the article) and this does not allow for modeling structural change in financial markets, with conditional volatility having different dynamics through time. It is in fact a misspecification error as the obtained coefficients have an upward bias after estimation.

For a study of this effect see[9]. First, let's have a measure of this persistence effect. If we consider the GARCH(1, 1) as the authors do in the introduction of their article, the sum $b_1 + c_1$ measures the persistence of a shock to the conditional variance in the equation of the conditional volatility. Given that this quantity must be strictly less than 1 (otherwise the variance is infinite when it reaches 1), the closer to 1 this quantity is, the more persistent is the process. Mikosch and Starica [10] first mention that for small samples this quantity can be significantly smaller than 1, but for larger ones it becomes close to 1. This phenomenon is refered to as Integrated-GARCH effect (where a IGARCH(1,1) is a 'GARCH(1,1)' with $b_1 + c_1 = 1$). An explanation proposed by the authors is that it can result from non-stationarity in larger data samples e.g here change in the unconditional variance. In such cases GARCH models are no longer valid and an extended models must be used. It can then be fruitful to use models in which the parameters are allowed to change over time to take into account change in volatility regime.

1.3 Different related models incorporating regime change

In order to tackle the previous limitation, researchers have proposed miscellaneous models with regime change, especially through Markov switching models. This makes it possible to have certain parameters for a certain period of the time series, and then other parameters for the next period as seen in the course. In particular a MS-GARCH is a GARCH whose parameters a, b_i, c_i are functions of a hidden markov chain, and therefore are allowed to change over time. They can be noted as $a(\Delta_t), b_i(\Delta_t), c_i(\Delta_t)$ where (Δ_t) is a finite state ergodic Markov chain independent from η.

However ML estimation of parameters is not feasible with MS-GARCH models, because one would need to integrate over all the possible regime paths, as it will be presented in 2.

To avoid this problem researchers have used ARCH parametrization, as in [8]. In this article, the authors also share the idea that there exist different regimes in financial time series. Indeed it seems reasonable to suggest that very large crisis have different consequences on the conditional volatility than small shocks. In the article, the authors quote Friedman and Laibson (1989) and Friedman (1992) who argued that conventional ARCH models fail to forecast well because large and small shocks have different effects. This difference in effects would then cause so-called 'high volatility' regimes and smaller ones.

To check the relevance of the introduction of regimes governed by an unobserved Markov Chain, they first fit a traditional ARCH. The fit is done on the one-week returns of the value-weighted portfolio of stocks traded on the New York Stock Exchange, ranging from 1962 to 1987. Then they fit an MS-ARCH (called SW-ARCH in their paper for 'SWitching') on the same data.

Their motivation was primarily that the forecast of traditional ARCH was quite poor, and that adding different regimes would give better performances. This is achieved by having different forecasts depending on the pasts regimes.

Indeed in the context of this model, the analyst can do different forecasts depending on the previous states. If the analyst is confident that the market has been in state 1 for the past periods (by smoothing, e.g recovering the past hidden states), that the probability p_{11} to stay in state 1 is close to unity, and if at date t the residual is small, then they would continue to place a high probability on the event that $s_t = 1$ (e.g the state at date t is 1), and the forecast would be that of the ARCH associated to state 1.

They conclude that such a specification offers both better fits and better forecasts. They also discovered that somehow the high-volatility regime is associated with economic recessions, which is in favor of what they imagined in the first place: shocks that have different amplitudes lead to different consequence, e.g different regimes.

However in their article they mainly focus on the forcasting performance of the MS-ARCH and not really on theoretical properties. The consistency of the MLE estimator of MS-ARCH has been established by Francq, Roussignol, and Zakoian (2001) [5]. In the article, the authors present necessary and sufficient conditions to ensure stationarity of the MS-GARCH and prove the consistency of the MLE for MS-ARCH. The MLE is derived after the parametric assumption that $\eta_t \sim N(0, 1)$ in the case where the martingale increment equation of 1.1 is written as $y_t = \sigma_t \eta_t$, but stated it can be extended to other distributions for η.

Other papers wanted to generalize to GARCH models by tackling the issue of the intractability of the MLE estimator by different means.

In [7], the authors wanted to mix several models together to model the short-term interest rate. They emphasized the difficulty to fit the data with many popular models, for instance stationarity is only guaranteed for special types of diffusion models, and estimates of GARCH models often imply an explosive conditional variance. However some of these models provide a good characterization for some given time range, even if after that they are no longer valid. The authors therefore wanted to nest many existing interest rate models as special cases and allow parameter values to change over time with change in regime.

In this context an MS-GARCH is introduced, with the big difference that it is a path-independant GARCH model. Each conditional variance depends only on the current regime, not on the whole past evolution of regimes and history of the process, which makes the likelihood easier to compute. At each point in time, dependence on the evolution of regimes can be "integrated out" by summing over all possible regimes to construct the variance conditional on observable information. For the interested reader, this is illustrated in the Fig. 2 of the article, where the authors present the dependence scheme using a tree. Indeed one can see that, by summing over all regimes in the central box of the tree, the dependence on past regimes goes away.

Finally, [3] can provide an estimation method based on a generalized method of moments (GMM) estimation. Indeed, in a previous paper [4] Christian Francq and Jean-Michel Zakoıan analyzed the L^2 structure of MS-GARCH processes and stated in their introduction that their results can be applied to identify the orders and to estimate the parameters of an MS-GARCH. This is what is done in [3].

To do so, the authors show that a MS-GARCH admits an ARMA representation in [4]. For instance, it is known that a GARCH model admits an ARMA representation by defining $u_t := y_t^2 - \sigma_t^2$ and $v_t := y_t^2 - \sigma^2$, the notations being that of 1.1, with in addition $\sigma^2 = E[\sigma_t^2] = cst$. One gets that $v_t \sim$ ARMA(p $\vee$ q, q).

Then in [3], they estimate, under some assumptions, the ARMA representations of y^2 and its powers thanks to the computation of autocorrelations, and recover the MS-GARCH parameters. In their first part, since MS-GARCH is a very large class of models, they show that sub-categories can be discriminated by looking at their autocovariance strucuture. Second, they go on with the estimation part. They illustrate the procedure by fitting a two-regime GARCH(1,1) model (that is shown to provide a reasonable fit) to the daily returns of the SP 500.

This part was about some of related modeling and estimation. It presented MS-ARCH and its very convenient estimation via Maximum Likelihood, and MS-GARCH with estimation via MLE in a particular case and via GMM. The next section is about the paper under-study.

2 Synthesis of the paper

The authors propose a Markov-switching GARCH model (MS-GARCH) to model the volatility of financial time series data, with the aim of addressing the issue of misspecification in standard GARCH models [1]. The MS-GARCH model allows the mean and variance of the process to switch in time between two different GARCH processes, governed by a hidden Markov chain. The authors provide sufficient conditions for the geometric ergodicity and the existence of moments of the process, and demonstrate that Bayesian estimation using a Gibbs sampling algorithm is feasible by enlarging the parameter space to include the state variables.

They also compare the performance of their model to a standard GARCH model using daily returns data for the SP500 index, and find that the MS-GARCH model is able to better capture the structural change in volatility present in the data.

2.1 The model

The GARCH(1,1) model is a statistical model that is used to analyze and predict time series data that exhibit volatility clustering, which means that the variance of the data tends to increase over time. The model is defined by two equations:

$$\begin{aligned} \mathrm{y}_t &= \mu_t \ + \ \sigma_t u_t \\ \sigma_t^2 &= \omega + \alpha\varepsilon_{t-1}^2 + \beta\sigma_{t-1}^2 \end{aligned} \tag{1}$$

the first part of equation (1) describes the relationship between the current value of the time series (y_t) and the mean and variance of the series at that point in time (μ_t and σ_t), and the second part of the equation describes the relationship between the variance at the current point in time (σ_t) and the variance at the previous point in time (σ_{t-1}). The model assumes that the error term (u_t) is independently and identically distributed with zero mean and unit variance. The parameters $\omega, \alpha,$ and β are non-negative (ω is strictly positive) and ensure that the conditional variance (σ_t) is positive.

The (MS-GARCH) model is a generalization of the GARCH(1,1) model that allows for different regimes for the conditional variance. The MS-GARCH model is defined by equation (2), which is similar to equations (1) of the GARCH(1,1) model, but the parameters $\omega_{s,t}, \alpha_{s,t},\ and\ \beta_{s,t}$ can vary across different regimes, which are defined by a finite set of discrete states (S) governed by a Markov chain with transition probabilities ($\eta_{i,j}$). The MS-GARCH model assumes that the error term (u_t) is independently and identically distributed with a continuous density function centered on zero and that the variance of the error term is finite. The model also assumes that the probabilities of remaining in each regime ($\eta_{i,i}$) are positive and that the expected value of the logarithm of the product of the conditional variance and the error term is negative across all regimes. These assumptions ensure that the process (y_t) defined by the MS-GARCH model is geometrically ergodic, and strictly stationary.

$$\begin{array}{ll} \mathrm{y}_t & = \mu_{s_t} \ + \ \sigma_t \\ \sigma_t^2 & = \omega_{s_t} + \alpha_{s_t}\varepsilon_{t-1}^2 + \beta_{s_t}\sigma_{t-1}^2 \end{array} \tag{2}$$

The geometric ergodicity is particularly important as it garantees the state space process converges to its stationary distribution at a geometric rate. A second important result is that under mild assumptions on the noise and the markov chain, y_t satisfies $\mathbf{E}_\pi[|y_t^2|^k] < \infty$ for some $k \geq 1$ where π is the stationary distribution of the state space process. This allows to say that the resulting MS-GARCH process is second-order stationary without necessarily each regime being so.

2.2 Estimation

The estimation of the model is done using bayesian inference and therefore by considering latent variables as parameters. Using a normality assumption, a two-regime model can be estimated by setting the set of parameters $\theta = (\theta_1', \theta_2')$ where $\theta_k' = (\omega_k, \alpha_k, \beta_k)$. The joint density of the complete paths of the state space process can then be written as:

$$f(y, S|\mu, \theta, \eta) \quad \propto \quad \prod_{t=1}^{T} \sigma_t^{-1} \exp\left(-\frac{(y_t - \mu_{s_t})^2}{2\sigma_t^2}\right) \eta_{s_t s_{t-1}} \tag{3}$$

Due to numerical infeasibility, authors use a Gibbs sampler to sample from posterior laws of parameters and states. For this purpose the authors specify the posterior densities of each parameter and use the following griddy-Gibbs sampler to sample from the parameters given the observations of the process:

- *Params* : Number of iterations R, A grid for each parameter (ex : $[\omega_1^1, ..., \omega_1^G]$ for ω_1)
- *Init* : chose parameter intializations (ex : $\omega_1^{(0)}$ for ω_1). This step can actually be quite problematic but intuitively one can chose the middle of the grid as a starting point even if this is not specified by the article)
- FOR r = 1,...,R:
  - FOR EACH PARAMETER IN $[\omega, \alpha, \beta, \mu, S]$:
    - compute conditional posterior cumulative distribution function at each point in the grid using deterministic integration over M points conditionally on the remaining parameters computed in the last iteration e.g compute:

$$f_i \quad = \quad \int_{\omega_1^1}^{\omega_1^i} \kappa(\omega_1|S^{(r)}, \beta_1^{(r)}, \alpha_1^{(r)}, \theta_2^{(r)}, \mu^{(r)}, y) d\omega_1 \quad , \quad i = 1, ..., G \tag{4}$$

      where κ is the kernel of the conditional posterior density of ω_1 given the remaining parameters and the process trajectory.

* sample from the posterior law using inverse cdf with numerical interpolation on the grid to sample next value of the parameter (the value at iteration r). This gives the value of the parameters at the current iteration (ex : $\omega^{(r)}$ for ω)

- *Output* : parameters at the each iteration.

The parameters sampled from the posterior laws before convergence of the sampler can then be discarded and the remaining observations can be used to approximate the posterior distribution of the model parameters.

2.3 Simulations and Application

The authors generated 50,000 observations from a data generating process supposing a standard normal noise, using prespecified parameters inspired by those of [8] where the transition probabilities out of each state are close to zero.

The process they use satisfies all the conditions for stationarity and existence of moments of high order. The data generating process observed seems to show an estimated density slightly skewed to the left and and the autocorrelation of squared data decreases slightly slower than the ACF of each component both almost at zero after lag 10. To reproduce these results a GARCH(1,1) process would have to be near integrated.

For the simulations the prior density of each parameter is set as a uniform between specified bounds. Simulations indicate that the posterior means are generally within one posterior standard deviation of the DGP values and there is no indication of label switching problems. Additionally, the authors found that by using a threshold of one-half to classify the states, 96% of the data were correctly classified. Such results show the empirical coherence and ability to reproduce observed facts of the model on simulated data.

The authors then continue and analyze the daily returns of the SP500 from 2001 to 2007, with a focus on the volatility of the returns. They describe the use of two different models to analyze the data, the MS-ARCH and MS-GARCH models, and compare their results.

The MS-ARCH model is found to have a weak evidence for a dynamic effect in the low volatility regime, while the MS-GARCH model is found to be more persistent and stable in the high volatility regime , and shows that the lagged conditional variance should be included in this regime. The differences in the means of the state variables between the two models suggest that the MS-ARCH model is less sensitive to high volatility periods. Additionally, unconditional probabilities of the regimes are estimated at 0.59 and 0.41 respectively, which is roughly consistent with the information provided by the mean states.

3 Critical analysis

The article under-study provides properties and an estimation on MS-GARCH models. However, some remarks can be made.

First, in terms of the employed method for the estimation and application to the SP500, the hidden Markov Chain has only two states and there is no justification why except 'for simplicity'. Indeed, it is a huge drawback of MS-GARCH models that the number of states can not be inferred in advance and even if there are methods such as AIC or BIC to infer the number of states, the good empirical performance of such criteria in this context is yet to be explored. Generally, the choice is made by trial and error.

To show the complexity of the choice of the number of states, we have decided to implement the DGP proposed by the paper by following the same parameters. The DGP is then fit using a Gibbs sampler in the R-package MSGARCH assuming there are 2 and 3 underlying states. The performance is then compared. It is important to note that in this implementation, the priors are taken to be standard normals. Figures 1,2 and 3 show respectively the trajectory for the 1500 first observations, the ACF of the squared data and the kernel density of the data.

In tables 2, 3 one can clearly see that the estimation precision is of the same order for the 3-state and the 2-state processes and that no significant estimation error in the 3-state process points to misspecification. This points out the difficulty of chosing the right specification for a MS-GARCH process on real data. Yet, one notices the two first states are very similar both in transition probabilities and in parameters which points out to a state being represented twice in the underlying state chain given that the underlying process is a two-state process. Although such remarks might slightly point towards a two-state model, table 1 shows that by using classical information criteria such as Deviance information criterion (DIC), a practitioner would not be able to decide between the two models and would even have a weak preference for the 3-state model (models with smaller DIC should be preferred to models with larger DIC).

Model	Deviance information criterion
3-state	165886.5966
2-state	165973.2639

Table 1: Deviance information criterion comparison for the two models

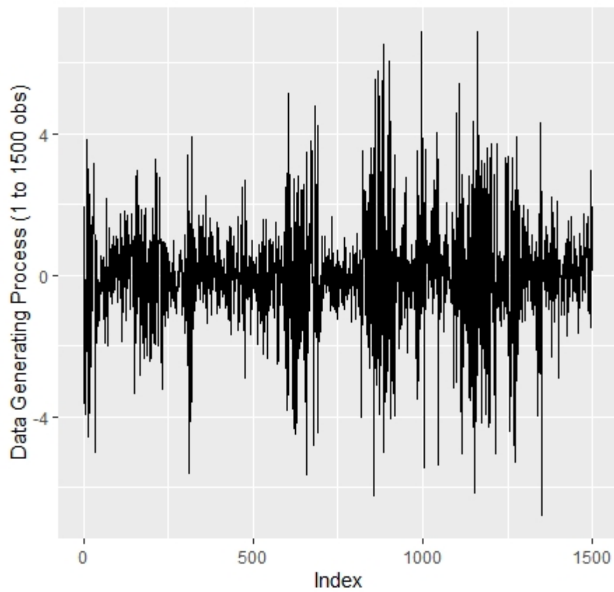

Figure 1: trajectory for the first 1500 observations

Variable	Mean	SD
ω_1	0.3196	0.0099
α_1	0.3323	0.0136
β_1	0.1994	0.0162
ω_2	3.9568	0.2736
α_2	0.1500	0.0124
β_2	0.2242	0.0343
$\eta_{1,1}$	0.9782	0.0012
$\eta_{2,1}$	0.0353	0.0017

Table 2: Parameter estimation for three state MS-GARCH on the DGP

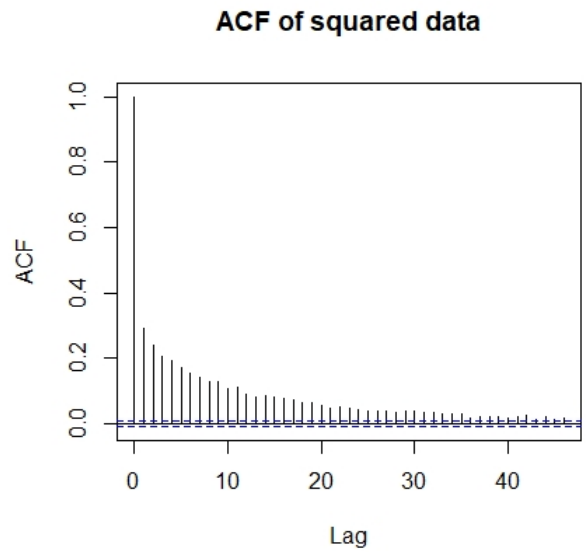

Figure 2: ACF of Squared DGP data

Variable	Mean	SD
ω_1	0.2938	0.0292
α_1	0.2365	0.0351
β_1	0.2309	0.0205
ω_2	0.3364	0.0236
α_2	0.3955	0.0319
β_2	0.1664	0.0223
ω_3	1.4917	0.1566
α_3	0.1110	0.0105
β_3	0.6583	0.0312
$\eta_{1,1}$	0.8459	0.0441
$\eta_{1,2}$	0.1396	0.0412
$\eta_{2,1}$	0.0583	0.0286
$\eta_{2,2}$	0.9194	0.0299
$\eta_{3,1}$	0.0179	0.0067
$\eta_{3,2}$	0.0173	0.0070

Table 3: Parameter estimation for three state MS-GARCH on the DGP

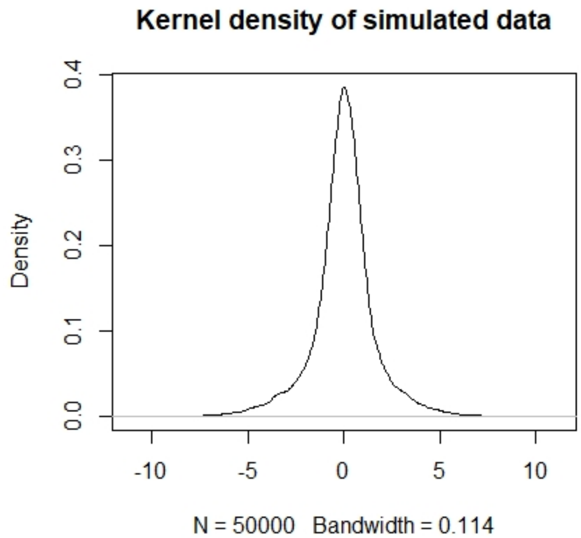

Figure 3: Kernel density of DGP Data

As regards the estimation of the article, it is a good thing that an estimation on a DGP (Data Generated process) was conducted. It makes it possible to see whether the estimation method (here a Gibbs) can actually recover the true parameters. However one can see that, even in the paper's estimation, the Gibbs sampler fails to recover some of the parameters especially α_1 (0.264 (0.051) instead of 0.35) and α_2 (0.142 (0.049) instead of 0.10). Moreover, it seems that the bounds of the priors are constrained to a rather small interval where the true parameter lies, which is unfortunate, since in practice one does not know the valid range for parameters. This constrained and informative prior therefore makes it easier for the algorithm to recover DGP parameters, but results may be different with flatter priors. In particular the sensitivity to the priors is not studied.

Additionally, in the example on the SP500, one could check whether the number of states is appropriate, as too many or too few states might lead to overfitting or underfitting of the data.

Finally a comparison of the results from this Bayesian estimation on the SP500 with those of [3] which uses a method based on GMM, or even to other related models such as a simple GARCH model (not only an MS-ARCH), and assessement of relative performance would have been welcome. Since the likelihood is not computed, this comparison cannot be done by comparing the log-likelihood values, but one could have looked at volatility forecasts for instance.

References

[1] Luc Bauwens, Arie Preminger, and Jeroen VK Rombouts. "Theory and inference for a Markov switching GARCH model". In: *The Econometrics Journal* 13.2 (2010), pp. 218–244.

[2] Tim Bollerslev. "Generalized autoregressive conditional heteroskedasticity". In: (1986).

[3] Jean-Michel Zakoıan Christian Francq. "Deriving the autocovariances of powers of Markov-switching GARCH models, with applications to statistical inference". In: (2007).

[4] Jean-Michel Zakoıan Christian Francq. "The L^2 structures of standard and switching-regime GARCH models". In: (2005).

[5] Michel Roussignol Christian Francq and Jean-Michel Zakoian. "Conditional heteroskedasticity driven by hidden markov chains". In: (1998).

[6] Engel. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation". In: (1982).

[7] Gray. "Modeling the conditional distribution of interest rates as a regime-switching process". In: (1996).

[8] James Hamilton and Raul Susmel. "Autoregressive conditional heteroskedasticity and changes in regime". In: *Journal of Econometrics* 64.1-2 (1994), pp. 307–333. URL: `https://EconPapers.repec.org/RePEc:eee:econom:v:64:y:1994:i:1-2:p:307-333`.

[9] Mikosch and Starica. "Non-stationarities in financial time series, the long range dependance and the IGARCH effect". In: (2004).

[10] Thomas Mikosch and Catalin Starica. "Long range dependence effects and ARCH modelling". In: (2004). URL: `https://EconPapers.repec.org/RePEc:wpa:wuwpem:0412004`.