ARTICLE IN PRESS

Computational Statistics and Data Analysis ▮ (▮▮▮▮) ▮▮▮–▮▮▮

Contents lists available at SciVerse ScienceDirect

Computational Statistics and Data Analysis

journal homepage: www.elsevier.com/locate/csda

# Maximum likelihood estimation of the Markov-switching GARCH model

Maciej Augustyniak *

*Département de mathématiques et de statistique, Université de Montréal, C.P. 6128, succursale Centre-ville, Montréal, Québec, Canada H3C 3J7*

ARTICLE INFO

*Article history:*
Received 16 March 2012
Received in revised form 19 December 2012
Accepted 29 January 2013
Available online xxxx

*Keywords:*
Markov-switching
GARCH
EM algorithm
Importance sampling

ABSTRACT

The Markov-switching GARCH model offers rich dynamics to model financial data. Estimating this path dependent model is a challenging task because exact computation of the likelihood is infeasible in practice. This difficulty led to estimation procedures either based on a simplification of the model or not dependent on the likelihood. There is no method available to obtain the maximum likelihood estimator without resorting to a modification of the model. A novel approach is developed based on both the Monte Carlo expectation–maximization algorithm and importance sampling to calculate the maximum likelihood estimator and asymptotic variance–covariance matrix of the Markov-switching GARCH model. Practical implementation of the proposed algorithm is discussed and its effectiveness is demonstrated in simulation and empirical studies.



## 1. Introduction

Financial time series exhibit complex statistical dynamics which are difficult to reproduce with stochastic models. These dynamics are often referred to as the stylized facts of financial data and include, among others, the heavy-tailed nature of the return distribution and volatility clustering (see Cont, 2001). The generalized autoregressive conditional heteroskedasticity (GARCH) class of models (Engle, 1982; Bollerslev, 1986) has been extensively used to model financial data as it offers an explicit way to model volatility. Markov-switching (MS) or regime-switching models have also attracted a lot of attention in the econometric literature since the seminal paper of Hamilton (1989). In MS models the return distribution at a given time depends on the state (or regime) of an unobserved Markov chain. The states of the Markov chain are often given an economic interpretation. For example, a regime with a negative mean return and high volatility may be associated with a state of financial distress in the economy.

Due to the popularity of MS and GARCH models, it is natural to combine these two approaches and consider a MS-GARCH model. The MS-GARCH model can be simply understood as a GARCH model where parameters depend on the state of an unobserved Markov chain. One way to justify such a combination is given by Lamoureux and Lastrapes (1990) and Mikosch and Starica (2004) who show that the high persistence observed in the variance of financial returns can be explained by time-varying GARCH parameters.

Hamilton and Susmel (1994) were among the first authors to discuss the MS-GARCH model. They noted that the estimation of this path dependent model is a challenging task because exact computation of the likelihood is infeasible in practice. This led some authors (Dueker, 1997; Gray, 1996; Haas et al., 2004; Klaassen, 2002) to propose estimating modified versions of the MS-GARCH model that circumvent the path dependence problem by maximum likelihood. Other authors suggested alternative estimation methods such as a generalized method of moments (GMM) procedure (Francq and

* Tel.: +1 514 5654726.
*E-mail address:* augusty@dms.umontreal.ca.
*URL:* http://www.dms.umontreal.ca/~augusty/.

0167-9473/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2013.01.026

Zakoïan, 2008) and a Bayesian Markov chain Monte Carlo (MCMC) algorithm (Bauwens et al., 2010, 2011). To this date, there is no method available to obtain the maximum likelihood estimator (MLE) of the MS-GARCH model without resorting to a simplification of the model.

The objective and main contribution of this article is to develop a novel approach based on the Monte Carlo expectation–maximization (MCEM) algorithm (Wei and Tanner, 1990) and the Monte Carlo maximum likelihood (MCML) method (Geyer, 1994, 1996) to estimate the MLE of the MS-GARCH model. The proposed algorithm requires simulations from the posterior distribution of the state vector. For this reason, it can be seen as a frequentist counterpart of the Bayesian MCMC method proposed by Bauwens et al. (2010) in the sense that both algorithms build on the data augmentation technique (Tanner and Wong, 1987). A secondary contribution of this article is to show how the asymptotic variance–covariance matrix of the MLE can be estimated. This is relevant since Francq and Zakoïan (2008) were not able to obtain the asymptotic standard errors of their GMM estimates due to numerical difficulties.

This paper is organized as follows. Section 2 defines the MS-GARCH model. Section 3 introduces the novel approach to calculate the MLE, proposes a procedure to approximate the asymptotic variance–covariance matrix of the MLE and discusses practical implementation of the algorithm. Section 4 demonstrates the effectiveness of the proposed method in a simulation study. Section 5 applies the estimation technique to daily and weekly log-returns on the S&P 500 index. Section 6 concludes and proposes avenues for further research. Moreover, Appendix A justifies the validity of the expectation–maximization (EM) algorithm when applied to the MS-GARCH model. Appendices B and C include a proof and some technical details related to the implementation of the algorithm.

## 2. The MS-GARCH model

### 2.1. Definition

Following Bauwens et al. (2010) and Francq et al. (2001), the MS-GARCH model can be defined by the following equations:

$$y_t = \mu_{S_t} + \sigma_t(S_{1:t})\eta_t, \tag{1}$$

$$\sigma_t^2(S_{1:t}) = \omega_{S_t} + \alpha_{S_t}\epsilon_{t-1}^2(S_{t-1}) + \beta_{S_t}\sigma_{t-1}^2(S_{1:t-1}), \tag{2}$$

$$\epsilon_{t-1}(S_{t-1}) = y_{t-1} - \mu_{S_{t-1}}. \tag{3}$$

The vector $(y_1, \ldots, y_T)$ represents the observations to be modeled and $\eta_t$, $t = 1, \ldots, T$, are independent and identically distributed normal innovations with zero mean and unit variance. At each time point, the conditional mean of the observation $y_t$ is $\mu_{S_t} = \mathrm{E}(y_t \mid S_t)$ and the conditional variance is $\sigma_t^2(S_{1:t}) = \mathrm{Var}(y_t \mid y_{1:t-1}, S_{1:t})$, where $y_{1:t-1}$ and $S_{1:t}$ are shorthand for the vectors $(y_1, \ldots, y_{t-1})$ and $(S_1, \ldots, S_t)$, respectively. The process $\{S_t\}$ is an unobserved ergodic time-homogeneous Markov chain with $N$-dimensional discrete state space (i.e., $S_t$ can take integer values from 1 to $N$). The $N \times N$ transition matrix of the Markov chain is defined by the transition probabilities $\{p_{ij} = \Pr(S_t = j \mid S_{t-1} = i)\}_{i,j=1}^N$. The vector $\theta = (\{\mu_i, \omega_i, \alpha_i, \beta_i\}_{i=1}^N, \{p_{ij}\}_{i,j=1}^N)$ denotes the parameters of the model. To ensure positivity of the variance, the following constraints are required: $\omega_i > 0$, $\alpha_i \geq 0$ and $\beta_i \geq 0$, $i = 1, \ldots, N$. Since $\sum_{j=1}^N p_{ij} = 1$ for $i = 1, \ldots, N$, $\theta$ contains $(4N + N(N-1))$ free parameters. Conditions for stationarity and the existence of moments were studied by Bauwens et al. (2010), Francq et al. (2001) and Francq and Zakoïan (2005).

### 2.2. Path dependence problem

The specification (1)–(3) causes difficulties in estimation since the conditional variance at time $t$ depends on the entire regime path $S_{1:t}$. To emphasize this dependence, the notation $\sigma_t^2(S_{1:t})$ is used in Eqs. (1)–(3), but to simplify it in what follows, $\sigma_t^2$ will be used to represent $\sigma_t^2(S_{1:t})$. Moreover, let $y$ and $S$ denote $y_{1:T}$ and $S_{1:T}$, respectively, and $f(p)$ stand for a probability density (mass) function. The calculation of the likelihood of the observations, denoted by $f(y \mid \theta)$, can be accomplished by integrating out all possible regime paths:

$$\begin{aligned} f(y \mid \theta) &= \sum_S f(y, S \mid \theta) = \sum_S f(y \mid S, \theta) p(S \mid \theta) \\ &= \sum_S \left[ \prod_{t=1}^T \sigma_t^{-1} \frac{1}{\sqrt{2\pi}} \exp\left( -\frac{(y_t - \mu_{S_t})^2}{2\sigma_t^2} \right) \right] p(S \mid \theta). \end{aligned} \tag{4}$$

For large $T$, this integration is infeasible numerically as the summation in Eq. (4) contains $N^T$ terms and quickly becomes very large. Even the estimation of the likelihood by brute force Monte Carlo (i.e., by simulating independent sequences of states from the underlying Markov chain) will fail since such estimators exhibit prohibitively large variances (see Danielsson and Richard, 1993). Nevertheless, as shown by Bauwens et al. (2011), it is possible to obtain an accurate estimate of the log-likelihood by writing

$$\log f(y \mid \theta) = \log f(y_1 \mid \theta) + \sum_{t=1}^{T-1} \log f(y_{t+1} \mid y_{1:t}, \theta),$$

Please cite this article in press as: Augustyniak, M., Maximum likelihood estimation of the Markov-switching GARCH model. Computational Statistics and Data Analysis (2013), doi:10.1016/j.csda.2013.01.026