HEC MONTREAL

Empirical and Simulation-Based Appraisal of a Class of Regime Switching GARCH Models

By

Monsede Franck Adoho

Department of Decision Sciences
(Financial Engineering)

In partial fulfillment of the requirements for the Degree of Master of Science (M.Sc.)

December 2019

Abstract

This thesis contributes to growing literature of Markov Regime Switching GARCH (RSG) models in two ways.

First, this thesis shows that the common path dependence problem of RSG models can be solved with the Expectation-Maximization Algorithm (EM). We simulate 100 data series with the path dependent RSG model with known parameters. We then use the EM algorithm to fit the model to the simulated data. Our results show that the EM algorithm infers the Data Generating Process (DGP) parameters satisfactorily and confirm the algorithm is able to estimate the path dependent Markov Chain Regime Switching GARCH models.

Second, with simulations studies, this thesis assess the empirical properties of the Markov Regime Switching GARCH (RSG) model of Gray (1996). We simulate 100 data series with the RSG model proposed by Gray (1996). We then fit the model with the simulated data using both the Maximum Likelihood Estimation (MLE) and the EM algorithm. Our results show that neither the ML algorithm nor the EM algorithm infers the DGP parameters correctly. These results confirm that model fails to identify the true parameters that generate the data. We conclude that the path independent Markov Chain Regime Switching GARCH model of Gray(1996) is not identified. The findings reinforce previous research from Haas, Mittnik and Paolla (2004) who raised concerns about the interpretability of the parameters of the RSG of Gray (1996).

The merite of this thesis is to be the first analytical work to show empirically (i) the limits of the Markov Regime Switching GARCH (RSG) model of Gray (1996) and (ii) to use the EM algorithm to fit the path dependent RSG model to data. We show that the EM is an alternative approach and perhaps an easier way of fitting the path dependent RSG models.

Key Words: Regime Switching, GARCH, Expectation Maximization, Maximum Likelihood, Mixture, Simulations.

Acknowledgement

I am grateful to Genevieve Gauthier for her availability and advice. I also want to thank my colleagues and friends for their help and support throughout this program.

To my family, I dedicate this thesis.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

In an effort to better understand the empirical regularities seen in the dynamics of financial assets' volatility, AutoRegressive Conditional Heteroskedasticity (ARCH) models and their generalized versions (GARCH) have been proposed respectively by Engle (1982) and Bollerslev (1986). These empirical regularities include high persistence and clustering of volatility, leverage effect, etc. GARCH models and its numerous extensions have been successfully used to model the dynamics of the volatility of various asset classes.

Like any parametric model, ARCH and GARCH models have their own limitations. In particular, it has been noted that GARCH models with constant parameters produce implausible results in certain applications (See e.g., Gray, 1996). This has led many authors to propose different formulations of Regime Switching GARCH (henceforth, RSG) models. Example of paper that study RSG models include Gray (1996), Haas, Mittnik, and Paolella (2004), Bauwens, Preminger and Rombouts (2010), Bauwens, Dufays and Rombouts (2014).

The traditional RSG models suffer from a path dependence problem that makes their estimation by maximum likelihood infeasible in practice. To solve this problem, two approaches have been used in the literature. The first approach inspired by Gray(1996) amounts to modifying the traditional RSG to render its Maximum Likelihood Estimation (MLE) possible; see (Duerke (1997), Klaassen(2002), Haas et al.(2004)). The second stream of the literature promotes the estimation of the original RSG models with different algorithms; see Das and Yoo (2004), Bauwens, Preminger and Rombouts (2010), Henneke, Rachev and Fabozzi (2011).

This paper studies the empirical properties of the Regime Switching GARCH (RSG) model proposed by Gray (1996) and proposes an estimation of the traditional RSG models with the Expectation-Maximization (EM) algorithm. The RSG model of Gray (1996) is appealing because it does not exhibit the path dependent that is typical to most RSG models. The method collapses the conditional variances into a single variance by computing their conditional expectation based on the regime probabilities. Gray (1996) specified this model with the intention of modeling the volatility of increments of an interest rate process. In that context, he argued that the RSG model is flexible enough to accommodated mean reversion in the interest rate process.

To begin, we use common random number simulations to 100 data series for the traditional path dependent RSG model and the RSG model of Gray (1996). To assess empirical properties of the (RSG) model proposed by Gray (1996), we fit the model with the simulated data using the Maximum Likelihood Estimation (MLE). Our results show that the RSG model of Gray (1996) does not infer the DGP parameters correctly. We conclude that the path independent Markov Chain Regime Switching GARCH model of Gray (1996) is not identified. Our findings echoes previous research by Haas, Mittnik and Paollella (2004) who also raised concerns about the consequences of recombination of the conditional variances.

We also investigate by simulations the ability of the EM algorithm to identify the DGP parameters of the path dependent Regime Switching GARCH (PD-RSG) model. We shows that the common path dependence problem of RSG models can be solved with the Expectation-Maximization Algorithm (EM). We use the EM algorithm to fit the model with simulated data. Our results confirm that the EM algorithm infers the Data Generating Process (DGP) parameters satisfactorily and we conclude that the algorithm is able to estimate the path dependent Markov Chain Regime Switching GARCH models.

The remainder of the paper is organized as follows. Section 2 reviews the GARCH and the Regime Switching GARCH models and their basic properties. Section 3 revisits the RSG model of Gray (1996). Section 4 presents the conditional mixture representation of the RSG model and the EM algorithm. Section 5 present the simulation study. Section 6 presents the empirical experiments and Section 7 concludes. The MATLAB codes used in the paper are presented in appendix.

Chapter 2

Traditional Markov Switching GARCH Models

2.1 GARCH Models

Let us consider a stationary process y_t. The basic ARCH(p) model proposed by Engle (1982) assumes that:

$$y_t = \mu_t + \sqrt{h_t}\varepsilon_t, \tag{2.1}$$

$$h_t = w + \sum_{i=1}^{p} a_i e_{t-i}^2. \tag{2.2}$$

where $e_t = y_t - \mu_t = \sqrt{h_t}\varepsilon_t$, ε_t is IID $N(0,1)$, $w > 0$, $a_1, ..., a_{p-1} \geq 0$ and $a_p > 0$. The mean process μ_t accommodates a wide range of specifications, including those that lead to view y_t as an ARMA process. Indeed, the ARCH denomination applies only to the specification of the volatility. In this thesis, y_t is the changes in the (logarithm of the gross) short interest rate. However, y_t can be something else in practice, for instance, the log-return on a financial asset.

ARCH models have been used with relative success in many empirically relevant situations. However, they may lack parsimony in cases where a large number of lagged error terms (p) is needed in order to fit the data correctly. In an effort to overcome this limitation, Bollerslev (1986) proposed GARCH(p,q) models, which parsimoniously extend ARCH(p) models to accommodate arbitrarily large values of p. A GARCH(p, q) model specifies the volatility process as:

$$h_t = w + \sum_{j=1}^{p} b_j h_{t-j} + \sum_{i=1}^{q} a_i e_{t-i}^2. \tag{2.3}$$

Andersen and Bollerslev (1998) argued that a standard GARCH(1,1) model provides accurate forecast of volatility in most cases of practical interest. Since then, it has become customary to think of volatility processes as GARCH(1,1), that is:

$$h_t = w + a e_{t-1}^2 + b h_{t-1}. \tag{2.4}$$

where $w > 0$, $a > 0$, $b > 0$ and $a + b < 1$.

The GARCH(1,1) model implies that:

$$Var\left(y_t | \Omega_{t-1}\right) = h_t, \tag{2.5}$$

where Ω_t is the information available up to time t. Also, Equation (2.4) implies that:

$$\begin{aligned} E\left(h_t\right) &= w + aE\left(e_{t-1}^2\right) + bE\left(h_{t-1}\right) \\ &= w + aE\left(h_{t-1}\varepsilon_{t-1}^2\right) + bE\left(h_{t-1}\right). \end{aligned}$$

By the Law of Iterated Expectation, we have:

$$\begin{aligned} E\left(h_{t-1}\varepsilon_{t-1}^2\right) &= E\left[E\left(h_{t-1}\varepsilon_{t-1}^2 | \Omega_{t-2}\right)\right] \\ &= E\left[h_{t-1}E\left(\varepsilon_{t-1}^2 | \Omega_{t-2}\right)\right] = E\left(h_{t-1}\right) \end{aligned}$$

By substituting into the previous equation and assuming stationarity for h_t, we obtain:

$$E(h_t) = \frac{w}{1-a-b}.$$

Finally, the law of total variance yields:

$$Var(y_t) = Var(\mu_t) + \frac{w}{1-a-b}. \tag{2.6}$$

This shows that the constraint $a + b < 1$ in necessary in order to have a stationary conditional variance process.

Although GARCH models are very popular because of their simplicity, easy of estimation and their empirical success in modeling time-varying volatility (Bollerslev, Chou,and kroner, 1992), an empirical common feature of these models is that, they tend to impute a high degree of persistence to conditional volatility. This means that shocks to the conditional variance which occurred in the distant past continue to have a non trivial impact in the current estimate of volatility. Engel and Bollerslev (1996) show that when $a + b$ is closed to one shocks to the conditional variance is hightly persistent and the model is possibly integrated. Several empirical studies (for example French et al., 1987; Chou , 1988; Fong, 1997) has reported values of $a + b$ above 0.9 for weekly stock returns. Lamoureux and lastrapes (1990) associate these high levels of volatility persistence with structural break in the volatility process. They point out that the persistence in variance may be overstated because of presence of, and failure to account for, deterministic structural shifts in the model. They, then show that GARCH measures of persistence in variance are sensitive to this type of model mis-specification. Diebold (1986) also indicated that the high persistence of volatility displayed by interest rate equations may be due to the failure to include a monetary-regime dummy for the conditional variance intercept.These empirical results, suggest that to obtain a more robust estimate of the conditional variance, a more general GARCH model that allows for regime shifts should be considered.

2.2 Markov Regime Switching GARCH Models

One popular way of making the GARCH models more flexible is allowing the parameters to change stochastically between regimes. The regimes are the outcomes of an unobserved Markov-chain process.

$$\begin{aligned} y_t &= \mu_{t,s_t} + \sqrt{h_{t,s_t}}\varepsilon_t, && (2.7)\\ h_{t,s_t} &= w_{s_t} + a_{s_t}e_{t-1}^2 + b_{s_t}h_{t-1}. && (2.8) \end{aligned}$$

where:

- $s_t \in \{1,2\}$ is a stationary and ergodic Markov Chain, with transition probabilities given by:
  $$\pi_= \begin{pmatrix} \pi_{11} & \pi_{21} \\ \pi_{12} & \pi_{22} \end{pmatrix},$$
  with π_{ij}=$Pr\left(s_t = j|s_{t-1} = i\right)$ and $\sum_j \pi_{ij} = 1$.
- $\mu_{t,s_t} = E\left(y_t|\Omega_{t-1}, s_t\right)$ is a state dependent conditional mean process;
- $h_{t,s_t} = Var\left(y_t|\Omega_{t-1}, s_t\right)$ is a state dependent conditional volatility process;
- ε_t is IID $N(0,1)$ and independent of h_{t,s_t};
- $e_t = y_t - E\left(y_t|\Omega_{t-1}\right)$ and $h_t = Var\left(y_t|\Omega_{t-1}\right)$, where $\Omega_t = \{y_t, ..., y_1\}$ is the information available at time t.
- $\left(w_i, a_i, b_i,\right), i = 1,2$, are regime specific parameters of the volatility process.

  The probability law that generates the observed data can be described as followed:
  $$y_t|\Omega_{t-1} = \begin{cases} N(\mu_{t,1}, h_{t,1}), & \text{w.p. } p_{t,1} \\ N(\mu_{t,2}, h_{t,2}), & \text{w.p. } p_{t,2} \end{cases} \qquad (2.9)$$

where $p_{t,i} = Pr(s_t = i|\Omega_{t-1})$ represents the state probability or the probability that given the information set of $t-1$ the observation y_t is drawn from the regime i. To fully describe the probability law governing the data, the conditional mean $\mu_{t,i}$, the conditional variance $h_{t,i}$ and the state probability $p_{t,i}$ must be specified.

Specification of the conditional mean

A general specification of the conditional mean is defined as follow:

$$\mu_{t,i} = \alpha_i + \beta_i X_{t-1}$$

where the explanatory variable which form the column of X_{t-1} are assumed to be exogenous in the sense that $\mathrm{E}[e_t|X_{t-1}]$=0. This means that the explanatory variable in the conditional mean equation are not independent from the error term.

Specification of the conditional variance

The conditional variance $h_{t,i}$ depends on the entire past history of the process.

$$h_{t,i} = w_i + a_i e_{t-1}^2 + b_i h_{t-1,i}$$

Specification of the posterior probabilities

The posterior probability $p_{t,1}$ is determined as follows:

$$\begin{aligned} p_{t,1} &= \Pr\left(s_t = 1|\Omega_{t-1}\right) \\ &= \Pr\left(s_t = 1|y_{t-1}, \Omega_{t-2}\right) = \frac{f\left(y_{t-1}, s_t = 1|\Omega_{t-2}\right)}{f\left(y_{t-1}|\Omega_{t-2}\right)}, \end{aligned}$$

where $f\left(y_{t-1}, s_t|\Omega_{t-2}\right)$ is the joint density of y_{t-1} and s_t conditional on Ω_{t-2}.

Hence, we have:

$$
\begin{aligned}
p_{t,1} &= \frac{\sum_{i=1}^{2} f\left(y_{t-1}, s_t = 1, s_{t-1} = i|\Omega_{t-2}\right)}{\sum_{i=1}^{2} f\left(y_{t-1}, s_{t-1} = i|\Omega_{t-2}\right)} \\
&= \frac{\sum_{i=1}^{2} f\left(y_{t-1}|s_t = 1, s_{t-1} = i, \Omega_{t-2}\right) \Pr\left(s_t = 1|s_{t-1} = i, \Omega_{t-2}\right) \Pr\left(s_{t-1} = i|\Omega_{t-2}\right)}{\sum_{i=1}^{2} f\left(y_{t-1}|s_{t-1} = i, \Omega_{t-2}\right) \Pr\left(s_{t-1} = i|\Omega_{t-2}\right)}.
\end{aligned}
$$

However, we note that:

$$
\begin{aligned}
f\left(y_{t-1}|s_t = j, s_{t-1} = i, \Omega_{t-2}\right) &= f\left(y_{t-1}|\Omega_{t-2}, s_{t-1} = i\right), \\
\Pr\left(s_t = j|s_{t-1} = i, \Omega_{t-2}\right) &= \Pr\left(s_t = j|s_{t-1} = i\right) = \pi_{ij} \text{ and} \\
f\left(y_t|\Omega_{t-1}, s_t = i\right) &= \frac{1}{\sqrt{2\pi h_{t,i}}} \exp\left(\frac{-\left(y_t - \mu_{t,i}\right)^2}{2h_{t,i}}\right) \text{ for all } t.
\end{aligned}
$$

Therefore, the dynamics of the posterior probabilities is described by the following recursion:

$$
\begin{aligned}
p_{t,1} &= \pi_{11} \frac{f\left(y_{t-1}|s_{t-1} = 1, \Omega_{t-2}\right) p_{t-1,1}}{\sum_{i=1}^{2} f\left(y_{t-1}|s_{t-1} = i, \Omega_{t-2}\right) p_{t-1,i}} \\
&\quad + \left(1 - \pi_{22}\right) \frac{f\left(y_{t-1}|s_{t-1} = 2, \Omega_{t-2}\right) p_{t-1,2}}{\sum_{i=1}^{2} f\left(y_{t-1}|s_{t-1} = i, \Omega_{t-2}\right) p_{t-1,i}}
\end{aligned} \tag{2.10}
$$

and $p_{t,2} = 1 - p_{t,1}$ for all t.

The Regime Switching Garch Model as defined in this section, (2.7) - (2.10), suffers from a path dependence problem that occurs because the conditional variance at time t depends on the entire sequence of regimes visited up to time t due to the recursive nature of the GARCH process. Even the GARCH model, (2.1)-(2.2) suffers from the path dependence problem, because h_t depends on its historical sequence up to time t although the couple h_{t+1} and y_t follows a Markov process. Subsequently, conditioning a regime switching GARCH model on a single regime would not solve the path dependence problem. To address the path dependence issue, Hamilton and Susmel (1994) and Cai (1994) applied a low-order ARCH process to model the conditional variance; which empirically would need many lags to capture the dynamic of the volatility. They justify the use of the ARCH process to model the conditional variance arguing that regime-switching GARCH models are intractable and impossible to estimate due the well known path dependency issue inherent to the GARCH process.

Gray (1996) shows that the problem of path dependence, can be solved in a way that preserves the essential nature of the GARCH process (including the important persistence effects) yet allows tractable estimation of the model. The next chapiter presents in detail the RSG proposed by Gray (1996).

Chapter 3

The Regime Switching GARCH Model of Gray (1996)

Gray (1996) argued that GARCH models with constant parameters produce *"untenable results"* when fitted to short interest rate processes. He therefore proposed to model the short rate as a Regime Switching GARCH (RSG) of the following form:

$$y_t = \mu_{t,s_t} + \sqrt{h_{t,s_t}}\varepsilon_t, \tag{3.1}$$

$$h_{t,s_t} = w_{s_t} + a_{s_t}e_{t-1}^2 + b_{s_t}h_{t-1}. \tag{3.2}$$

where:

- $s_t \in \{1,2\}$ is a stationary and ergodic Markov Chain, with transition probabilities given by:

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{21} \\ \pi_{12} & \pi_{22} \end{pmatrix},$$

with $\pi_{ij} = \Pr(s_t = j | s_{t-1} = i)$.

- $\mu_{t,s_t} = E\left(y_t|\Omega_{t-1}, s_t\right)$ is a state dependent conditional mean process;
- $h_{t,s_t} = Var\left(y_t|\Omega_{t-1}, s_t\right)$ is a state dependent conditional volatility process;
- ε_t is IID $N(0,1)$ and independent of h_{t,s_t};
- $e_t = y_t - E\left(y_t|\Omega_{t-1}\right)$ and $h_t = Var\left(y_t|\Omega_{t-1}\right)$, where $\Omega_t = \{y_t, ..., y_1\}$ is the information available to the econometrician at time t.
- $\left(w_i, a_i, b_i,\right), i = 1, 2$, are regime specific parameters of the volatility process.

In Gray (1996), y_t is assumed to be the first difference of a mean-reverting short interest rate process, which we denote r_t. Mean reversion is accounted for by letting $\mu_{t,s_t} = \alpha_{s_t} + \beta_{s_t} r_{t-1}$ with $\beta_{s_t} < 0$. In other applications, μ_{t,s_t} may be specified as constant per regime ($\mu_{t,s_t} \in \{\mu_1, \mu_2\}$) or as function of exogenous variables ($\mu_{t,s_t} = \alpha_{s_t} + \beta_{s_t} X_{t-1}$). Model (3.1)-(3.2) is appealing because it avoids the path dependence problem akin to regime switching volatility models. Indeed, the motion of h_{t,s_t} is specified in terms of h_{t-1} regardless of the current state. In a path dependent model, the dynamics of the conditional variance process would be specified as:

$$h_t = w_{s_t} + a_{s_t} e_{t-1}^2 + b_{s_t} h_{t-1}. \tag{3.3}$$

where h_t now denotes the variance of y_t conditional on Ω_t and $\{s_t, ..., s_1\}$. In this case, the computation of the likelihood of a sample of size T involves an integration over 2^T possible paths of the latent Markov Chain s_t. This curse of dimensionality is avoided by Gray's formulation.

Marginalizing s_t out of the conditional mean in Gray's model yields:

$$\begin{aligned} E\left(y_t|\Omega_{t-1}\right) &= p_{t,1} E\left(y_t|\Omega_{t-1}, s_t = 1\right) + \left(1 - p_{t,1}\right) E\left(y_t|\Omega_{t-1}, s_t = 2\right) \\ &= p_{t,1}\mu_{t,1} + \left(1 - p_{t,1}\right)\mu_{t,2} \equiv \mu_t, \end{aligned}$$

where $p_{t,1} = \Pr(s_t = 1|\Omega_{t-1})$. For the conditional variance, we have:

$$\begin{aligned} Var(y_t|\Omega_{t-1}) &= E\left(y_t^2|\Omega_{t-1}\right) - E(y_t|\Omega_{t-1})^2 \\ &= \sum_{i=1}^{2} p_{i,t}\left[h_{t,i} + \mu_{t,i}^2\right] - \left(\sum_{i=1}^{2} p_{i,t}\mu_{t,i}\right)^2 \equiv h_t. \end{aligned}$$

The volatility of the current period given a particular state s_t depends on the volatility that was expected at the previous period (h_t) unconditionally of the states. As the *"realized"* value of the volatility at the previous period can be higher or lower than the expected volatility, Gray's RSG model might be less good at replicating volatility clustering than a path dependent RSG model.

The posterior probability $p_{t,1}$ is determined as follows:

$$\begin{aligned} p_{t,1} &= \Pr(s_t = 1|\Omega_{t-1}) \\ &= \Pr(s_t = 1|y_{t-1}, \Omega_{t-2}) = \frac{f(y_{t-1}, s_t = 1|\Omega_{t-2})}{f(y_{t-1}|\Omega_{t-2})}, \end{aligned}$$

where $f(y_{t-1}, s_t|\Omega_{t-2})$ is the joint density of y_{t-1} and s_t conditional on Ω_{t-2}. Hence, we have:

$$\begin{aligned} p_{t,1} &= \frac{\sum_{i=1}^{2} f(y_{t-1}, s_t = 1, s_{t-1} = i|\Omega_{t-2})}{\sum_{i=1}^{2} f(y_{t-1}, s_{t-1} = i|\Omega_{t-2})} \\ &= \frac{\sum_{i=1}^{2} f(y_{t-1}|s_t = 1, s_{t-1} = i, \Omega_{t-2}) \Pr(s_t = 1|s_{t-1} = i, \Omega_{t-2}) \Pr(s_{t-1} = i|\Omega_{t-2})}{\sum_{i=1}^{2} f(y_{t-1}|s_{t-1} = i, \Omega_{t-2}) \Pr(s_{t-1} = i|\Omega_{t-2})}. \end{aligned}$$

However, we note that, in the regime switching GARCH,

$$\begin{aligned} f(y_{t-1}|s_t = j, s_{t-1} = i, \Omega_{t-2}) &= f(y_{t-1}|\Omega_{t-2}, s_{t-1} = i), \\ \Pr(s_t = j|s_{t-1} = i, \Omega_{t-2}) &= \Pr(s_t = j|s_{t-1} = i) = \pi_{ij} \text{ and} \\ f(y_t|\Omega_{t-1}, s_t = i) &= \frac{1}{\sqrt{2\pi h_{t,i}}} \exp\left(\frac{-(y_t - \mu_{t,i})^2}{2h_{t,i}}\right) \text{ for all } t. \end{aligned}$$

Therefore, the dynamics of the posterior probabilities is described by the following recursion:

$$\begin{aligned} p_{t,1} &= \pi_{11} \frac{f\left(y_{t-1}|s_{t-1}=1,\Omega_{t-2}\right)p_{t-1,1}}{\sum_{i=1}^{2} f\left(y_{t-1}|s_{t-1}=i,\Omega_{t-2}\right)p_{t-1,i}} \\ &\quad + (1-\pi_{22}) \frac{f\left(y_{t-1}|s_{t-1}=2,\Omega_{t-2}\right)p_{t-1,2}}{\sum_{i=1}^{2} f\left(y_{t-1}|s_{t-1}=i,\Omega_{t-2}\right)p_{t-1,i}} \end{aligned} \tag{3.4}$$

and $p_{t,2} = 1 - p_{t,1}$ for all t.

The RSG model proposed by Gray (1996) departs from the traditional RSG model in two ways. At any given time t, the conditional mean μ_t is calculated as the expectation of μ_{t,s_t} over the two regimes and a single variance h_t is computed by recombining the conditional variances over the two regimes. This method collapses the conditional variances in each regime by taking the conditional expectation of h_t based on the regime probabilities. The path dependency is removed as h_t, the conditional variance at time t depends only on the current regime and the information set available at $t-1$ while the GARCH effect are still allowed. Klaassen (2002) noted that an important inconvenient to Gray's model is its inability to generating multi-period variance forecast and suggested a modified version of the model.

Chapter 4

Fitting the RSG Model to Data

As discussed at the end of the chapter 2 and chapter 3 the path dependent RSG and the Gray's family RSG model can not be solved analytically. In fact, the computation of the likelihood function of the path dependent RSG is infeasible in practice. Although the Gray's model solve the path dependency issue and hence can be estimated by a maximum likelihood method, it displays serious drawback. Haas, Mittnik and Paolella (2004) argued that by collapsing the two conditional variances into a single variance, the economic significance of the variance dynamics, in Gray's model, became unclear and, the disaggregation of the overall variance process provided by the model was at best difficult to interpret. More generally, Haas, Mittnik and Paolella (2004) noted that many models that combine GARCH features with regimes shifts "*suffer from severe estimation difficulties*" and "*their dynamic properties are not well understood.*" Several versions of the Regime Switching Garch models that can be estimated by a maximum likelihood method have been proposed; see (Duerke (1997), Klaassen(2002), Haas et al.(2004)). Rather than modifying the original path dependent RSG, other researched have proposed a Bayesian Markov Chain Monte Carlo (MCMC) algorithm that overcomes the path dependence; see Das and Yoo (2004), Bauwens, Preminger and Rombouts (2010), Henneke, Rachev and Fabozzi (2011). This thesis shows that the path dependence problem can be solved with the Expectation-Maximization Algorithm. In this section, we present two strategies to estimate the parameters of the path dependent RSG model and the RSG model of Gray (1996). The first approach relies on the likelihood estimation of a conditional mixture representation of the RSG model and the second approaches is based on an Expectation-Maximization (EM) algorithm of the same. This thesis is the first to use the EM algorithm to test the empirical properties of the path dependent RSG model of and the RSG model of Gray (1996).

4.1 Maximum Likelihood (ML)

We use the notation $f\left(y_t|\Omega_{t-1};\theta\right)$ instead of $f\left(y_t|\Omega_{t-1}\right)$ to highlight the dependence of the likelihood function on θ, where θ is a vector that collects all parameters of the model. The sample likelihood function of the RSG model can be factorized as:

$$\begin{aligned} L\left(\theta\right) &= f\left(\Omega_t;\theta\right) \\ &= f\left(y_1;\theta\right)\prod_{t=2}^{T} f\left(y_t|\Omega_{t-1};\theta\right), \end{aligned}$$

where $f\left(y_1;\theta\right)$ is the marginal likelihood of the RSG process. Therefore, the sample log-likelihood is:

$$L\left(\theta\right) = \sum_{t=2}^{T} \log f\left(y_t|\Omega_{t-1};\theta\right) + \log f\left(y_1;\theta\right).$$

Unfortunately, the exact form of the marginal likelihood $f\left(y_1;\theta\right)$ is unknown. For simplicity, we have removed the term $\log f\left(y_1;\theta\right)$ from the sample log-likelihood. This amounts to say that the parameters are estimated conditional on the first observation.

By noting that $f\left(y_t|\Omega_{t-1};\theta\right)$ is obtained by marginalizing the joint $f\left(y_t, s_t = i|\Omega_{t-1};\theta\right)$

with respect to s_t, we have:

$$\begin{aligned} L(\theta) &= \sum_{t=2}^{T} \log\left(\sum_{i=1}^{2} f\left(y_t, s_t = i | \Omega_{t-1}; \theta\right)\right) \qquad (4.1) \\ &= \sum_{t=2}^{T} \log\left[p_{t,1} f\left(y_t | \Omega_{t-1}, s_t = 1; \theta\right) + p_{t,2} f\left(y_t | \Omega_{t-1}, s_t = 2; \theta\right)\right]. \end{aligned}$$

Finally, the Maximum Likelihood (ML) estimator is obtained by maximizing $L(\theta)$ with respect to θ.

There are well-known numerical problems associated with the maximization of the likelihood of mixtures. Indeed, $L(\theta)$ is not necessarily globally convex in θ due to the summation inside the logarithm. Moreover, the sample likelihood of mixture models can be unbounded. This happens when one of the regimes collapses to only a few observations so that the variance under that regime is close to zero. Nevertheless, we will estimate the RSG model based on the likelihood function above and compared the results with those of the EM algorithm presented below.

4.2 Expectation Maximization (EM)

The Expectation-Maximization (EM) algorithm has been proposed in an effort to avoid the numerical difficulties associated with the conditional mixture approach. This technique due to Dempster, Laird and Rubin (1977) updates the parameter θ iteratively along two steps. The first step (E-step) evaluates the posterior probabilities at the current value of the parameter and uses them as input in an auxiliary objective function. The second step (M-step) computes the updated value of the parameters by maximizing the auxiliary objective function.

Let θ_{m-1} denote the current value of the parameter. To build our auxiliary objective function, we note that:

$$\begin{aligned} L(\theta) - L(\theta_{m-1}) &= \sum_{t=2}^{T} \log \frac{f(y_t, |\Omega_{t-1}; \theta)}{f(y_t|\Omega_{t-1}; \theta_{m-1})} \\ &= \sum_{t=2}^{T} \log \sum_{i=1}^{2} \frac{f(y_t, s_t = i|\Omega_{t-1}; \theta)}{f(y_t|\Omega_{t-1}; \theta_{m-1})} \frac{\Pr(s_t = i|\Omega_{t-1}; \theta_{m-1})}{\Pr(s_t = i|\Omega_{t-1}; \theta_{m-1})} \end{aligned}$$

Jensen's inequality implies:

$$\begin{aligned} & L(\theta) - L(\theta_{m-1}) \\ \geq\ & \sum_{t=2}^{T} \sum_{i=1}^{2} \Pr(s_t = i|\Omega_{t-1}; \theta_{m-1}) \log \frac{f(y_t, s_t = i|\Omega_{t-1}; \theta)}{f(y_t|\Omega_{t-1}; \theta_{m-1}) \Pr(s_t = i|\Omega_{t-1}; \theta_{m-1})} \\ =\ & \sum_{t=2}^{T} \sum_{i=1}^{2} \Pr(s_t = i|\Omega_{t-1}; \theta_{m-1}) \log f(y_t, s_t = i|\Omega_{t-1}; \theta) \\ & - \sum_{t=2}^{T} \sum_{i=1}^{2} \Pr(s_t = i|\Omega_{t-1}; \theta_{m-1}) \log \left[f(y_t|\Omega_{t-1}; \theta_{m-1}) \Pr(s_t = i|\Omega_{t-1}; \theta_{m-1}) \right] \end{aligned}$$

The second double summation does not depend on θ. Therefore, our auxiliary objective function is taken to be:

$$Q(\theta, \theta_{m-1}) = \sum_{t=2}^{T} \sum_{i=1}^{2} \Pr(s_t = i|\Omega_{t-1}; \theta_{m-1}) \log f(y_t, s_t = i|\Omega_{t-1}; \theta), \tag{4.2}$$

where:

$$\log f(y_t, s_t = i|\Omega_{t-1}; \theta) = \log f(y_t|\Omega_{t-1}, s_t = i; \theta) + \log \Pr(s_t = i|\Omega_{t-1}; \theta).$$

Therefore, the EM algorithm proceeds as follows, after setting an initial value θ_0:

- **E-step:** Use the current value of the parameter to compute the posterior probabilities $\Pr(s_t = i|\Omega_{t-1}; \theta_{m-1})$, $t = 1, ..., T$ and $i = 1, 2$.

- **M-step:** Update the value of the parameter by solving:

$$\theta_m = \arg\max_{\theta} \left(Q\left(\theta, \theta_{m-1}\right)\right). \tag{4.3}$$

These two steps are iterated until convergence, e.g., until $\|\theta_m - \theta_{m-1}\|$ becomes sufficiently small. It has be shown in standard cases that the EM algorithm converges in a finite number of steps, possibly to a local maximum (see Navidi, 1997).[1]

After obtaining an estimate $\widehat{\theta}$ of θ, the updated posterior probabilities of the **E-step**, $\Pr(s_t = i|\Omega_t; \theta_{m-1})$ may be calculated as follows:

$$\Pr\left(s_t = i|\Omega_t; \widehat{\theta}\right) = \frac{f\left(y_t|\Omega_{t-1}, s_t = i; \widehat{\theta}\right) p_{t,i}\left(\widehat{\theta}\right)}{\sum_{i=1}^{2} f\left(y_t|\Omega_{t-1}, s_t = i; \widehat{\theta}\right) p_{t,i}\left(\widehat{\theta}\right)} \tag{4.4}$$

where $p_{t,i}(\theta) = \Pr(s_t = i|\Omega_{t-1}; \theta)$ for $i = 1, 2$. The updated posterior probabilities, $\Pr(s_t = i|\Omega_t; \theta)$, are supposed to be more informative about the actual state prevailing at time t than the filtered probabilities, $\Pr(s_t = i|\Omega_{t-1}; \theta)$.

[1] The EM algorithm does not necessarily converge to the global maximum. This risk can be minimized by initializing the algorithm at several differents initial points and comparing the values of the likelihood function at convergence.

Chapter 5

Simulation studies for model validation: case of Path dependent RSG model

In this chapter, we simulated a data-generating process (DGP) corresponding to the path dependent model presented in Chapter 2. We simulated 100 series of 10000 weekly observations to which we fit the RSG model to assess its ability to infer the parameters used to generate the data.

5.1 Simulation of the Path Dependent Markov Chain RSG Model

In this section, we study by simulation the performance of the full path dependent RS-Garch model (2.7 - 2.10) discussed in chapter 2. The first subsection presents the design of the simulation, the second subsection presents the algorithm used to evaluate the objective functions and the third subsection presents the simulations results.

5.1.1 Simulation Strategy and Design

We simulate $M = 100$ samples of size $T = 10000$ from the model (2.7), (2.8), (2.9) and (2.11) using the following parameter values :

$$
\begin{aligned}
\alpha_1 &= 0.25,\ \beta_1 = 0.0008,\ w_1 = 0.5,\ a_1 = 0.35,\ b_1 = 0.4, \\
\alpha_2 &= -0.07,\ \beta_1 = -0.0006,\ w_2 = 0.08,\ a_2 = 0.1,\ b_2 = 0.2.
\end{aligned}
$$

Note that the parameters have been chosen for illustration purpose; the Regime 1 displays positive returns and high volatility and Regime 2 displays low volatility and low return. The long term conditional volatilities associated with these parameters are:

$$
\begin{aligned}
v_1 &= \frac{0.5}{1 - 0.35 - 0.4} = 2 \\
v_2 &= \frac{0.08}{1 - 0.1 - 0.2} = 0.11
\end{aligned}
$$

Moreover, Regime 1 displays positive returns as the parameters of the conditional mean are all positive while in Regime 2 the conditional mean parameters are all negative suggesting low returns compared to Regime 1.

The probability of transitions of s_t are:

$$
\Pi = \begin{pmatrix} \pi_{11} & \pi_{21} \\ \pi_{12} & \pi_{22} \end{pmatrix} = \begin{pmatrix} 0.9739 & 0.010\,4 \\ 0.026\,1 & 0.9896 \end{pmatrix}.
$$

The parameters above suggest that the Markov chain has very persistent regimes while the volatility process in each regime is not very persistent. The ergodic probabilities associated with these transition probabilities are:

$$
\begin{aligned}
\pi_1 &= \frac{1 - 0.9896}{2 - 0.9739 - 0.9896} = 0.285 \\
\pi_2 &= \frac{1 - 0.9739}{2 - 0.9739 - 0.9896} = 0.715
\end{aligned}
$$

Hence, Regime 2 is more than twice as probable as Regime 1 in the stationary distribution. In all the simulation exercise, we use the same variables as in Gray's paper. In fact, $y_t = r_t - r_{t-1}$ represented a serie of returns of interest rate r_t and $X_{t-1}= r_{t-1}$. Hence, we set the initial values of the variables as follow:

$$
\begin{aligned}
s_0 &= 1,\, y_0 = 0,\, r_0 = 5 \text{ and} \\
h_{0,i} &= Var\,(y_0|\text{no information}) = \frac{w_i}{1 - a_i - b_i}.
\end{aligned}
$$

where $r_t = r_0 + \sum_{k=1}^{t} y_t$ is the cumulative process; $r_t = r_{t-1} + y_t$.

$$
f_{1,i} = \frac{1}{\sqrt{2\pi h_{0,i}}} \exp\left(-\frac{y_1^2}{2h_{0,i}}\right) = \frac{1}{\sqrt{2\pi h_{0,i}}}, \text{ for } i = 1, 2.
$$

The posterior probabilities are initialized to the ergodic probabilities:

$$
p_{1,1} = \pi_1 = 0.285 \text{ and } p_{1,2} = \pi_2 = 0.715
$$

Next, we apply a recursion for $t = 2$ to T. At a given step t, we go through the following sub-steps:

- Step 1: Compute the potential values of the volatility as:

$$
\begin{aligned}
h_{t,1} &= w_1 + a_1 e_{t-1}^2 + b_1 h_{t-1,1} \\
h_{t,2} &= w_2 + a_2 e_{t-1}^2 + b_2 h_{t-1,2}
\end{aligned}
$$

with the convention that $e_0 = 0$. One of $h_{t,1}$ and $h_{t,2}$ is the realized value of the volatility at time t and both are needed to update h_{t-1}.

- Step 2: Draw a random number U_t from the uniform distribution on $[0, 1]$. If $s_{t-1} = 1$, then set $s_t = 1$ if $U_t \leq \pi_{11}$ and $s_t = 2$ otherwise. If $s_{t-1} = 2$, then set $s_t = 2$ if $U_t \leq \pi_{22}$ and $s_t = 1$ otherwise.

- Step 3: Compute $y_t = \mu_t^* + \sqrt{h_t^*}\varepsilon_t$, where ε_t is an independent draw from $N(0, 1)$, $\mu_t^* = \alpha_{s_t} + \beta_{s_t} r_{t-1}$, $h_t^* = h_{t,s_t}$. Likewise, compute the cumulative process as $r_t = r_0 + \sum_{k=0}^{t} y_t$ for all t.

- Step 4: Update the posterior probabilities:

$$
\begin{aligned}
p_{t,1} &= \pi_{11}\frac{f_{t-1,1}p_{t-1,1}}{f_{t-1,1}p_{t-1,1} + f_{t-1,2}p_{t-1,2}} + (1-\pi_{22})\frac{f_{t-1,2}p_{t-1,2}}{f_{t-1,1}p_{t-1,1} + f_{t-1,2}p_{t-1,2}} \\
p_{t,2} &= 1 - p_{t,1}
\end{aligned}
$$

where $f_{t,i} = f\left(r_t | s_t = i, \Omega_{t-1}\right)$ for all t.

- Step 5: Update the likelihoods:

$$
f_{t,i} = \frac{1}{\sqrt{2\pi h_{t,i}}}\exp\left(-\frac{y_t^2}{2h_{t,i}}\right), i = 1, 2
$$

Note that **Step 1 to 3:** are sufficient for data simulation. The next subsection explains how to evaluate the objective functions.

5.1.2 Evaluating the Objective Functions

Both the sample log-likelihood of the mixture representation $L(\theta)$ given at (4.1) and the auxiliary objective function of the EM algorithm $Q(\theta, \theta_{m-1})$ given at (4.2) depend on the posterior probabilities $p_{t,1} = \Pr(s_t = i|\Omega_{t-1}; \theta)$ and $p_{t,2} = 1 - p_{t,1}$. Indeed, $L(\theta)$ is a sum of logarithms of a weighted average of two likelihood functions while $Q(\theta, \theta_{m-1})$ is a sum of weighted average of log-likelihoods.

In order to evaluate $L(\theta)$ and $Q(\theta, \theta_{m-1})$, we design two separate functions:

- a MATLAB function (called *ProbRsGarchPD.m*) that takes θ_{m-1} and $\{y_t, r_{t-1}\}_{t=1}^{T}$ as input and returns $p_{t,i}(\theta_{m-1})$, $f(y_t|\Omega_{t-1}, s_t = i)$, for $i = 1, 2$ and $t = 1, ..., T$;
- a MATLAB function (called *EM_RsGarchPD.m*) that takes $p_{t,i}(\theta_{m-1})$, θ and $\{y_t, r_{t-1}\}_{t=1}^{T}$ as input and returns $Q(\theta, \theta_{m-1})$.

The second function becomes trivial once the posterior probabilities $p_{t,i}(\theta_{m-1})$ are available. The presentation below therefore focuses on the first function, *ProbRsGarchPD.m*.

To begin, we initialize the posterior probabilities at the ergodic probabilities:

$$\begin{aligned} p_{1,1} &\equiv \frac{1-\pi_{22}}{2-\pi_{11}-\pi_{22}}, \\ p_{1,2} &\equiv \frac{1-\pi_{11}}{2-\pi_{11}-\pi_{22}}, \end{aligned}$$

where Ω_0 is an empty set. We have also set the initial conditional variances in each regime to be equal to the sample variance:

$$h_{1,i} \equiv \frac{1}{T-1} \sum_{t=1}^{T} (y_t - \overline{y})^2$$

for $i = 1, 2$, where $\overline{y} = \frac{1}{T-1} \sum_{t=1}^{T} y_t$.

Based on these initial values, we compute the first conditional likelihoods:

$$f(y_1|\Omega_0, s_1 = i) \equiv \frac{1}{\sqrt{2\pi h_{1,i}}} \exp\left(\frac{-(y_1 - \mu_{1,i})^2}{2h_{1,i}} \right).$$

where $\mu_{t,i} = \alpha_i + \beta_i r_{t-1}$ and $e_{t,i} = y_t\text{-}\mu_{t,i}$ for all i and t.

This completes the initialization step.

Next, we interactively compute the following quantities in a loop starting from $t = 2$ and ending at $t = T$:

- Conditional variance per regime:

$$h_{t,i} \equiv w_i + a_i e_{t-1}^2 + b_i h_{t-1,i};\ \ i = 1, 2.$$

- Likelihood per regime:

$$f\left(y_t|\Omega_0, s_t = i\right) \equiv \frac{1}{\sqrt{2\pi h_{t,i}}} \exp\left(\frac{-\left(y_t - \mu_{t,i}\right)^2}{2h_{t,i}}\right).$$

- Posterior probabilities

$$\begin{aligned} p_{t,1} &\equiv \pi_{11} \frac{f\left(y_{t-1}|s_{t-1} = 1, \Omega_{t-2}\right) p_{t-1,1}}{\sum_{i=1}^{2} f\left(y_{t-1}|s_{t-1} = i, \Omega_{t-2}\right) p_{t-1,i}} \\ &\quad + \left(1 - \pi_{22}\right) \frac{f\left(y_{t-1}|s_{t-1} = 2, \Omega_{t-2}\right) p_{t-1,2}}{\sum_{i=1}^{2} f\left(y_{t-1}|s_{t-1} = i, \Omega_{t-2}\right) p_{t-1,i}}, \\ p_{t,2} &\equiv 1 - p_{t,1}. \end{aligned}$$

At the end of this loop, we have all the inputs needed to evaluate the likelihood of the conditional mixture ($L(\theta)$) as well as the auxiliary objective function of the EM algorithm ($Q(\theta, \theta_{m-1})$). The codes that evaluate these objective functions are provided in Appendix.

5.2 Simulation Results

We use the data generated (100 samples of 10000 observations) following the simulation strategy presented in section 5.2.1 to fit the path dependent RSG model. The parameters value of the Data Generating Process suggest that the GARCH equation of the regime 1 is twice as persistent as the regime 2.

The transition probabilities are close to unity and for each conditional variance equation $a_i + b_i < 1$.

To confirm that the simulated data are drawn from different GARCH processes, we fit GARCH (1,1) to the square of the error term $(y_t - \mu_t)$ for each serie. The sum of the estimated parameters $\hat{a_i} + \hat{b_i} = 1$ for all 100 samples. These results suggest that only an integrated GARCH can fit each data serie, confirming a regime shift in the GARCH equations.

Next we estimate the parameters value of the simulated data using the path dependent RSG model (2.7 - 2.10). In Table 5.1, we report few statistics including the minimum, the maximum, the sample mean, the quartiles and the sample standard deviations of the estimated parameters. First, the sample mean of the estimated parameters are all close to the DGP value. Second, all DGP values fall withing the minimum and the maximum bounds of the estimated parameters. Third, we test the null hypothesis that the estimated parameters are equal to their DGP value and report the probability of rejecting (H0 Rejection Rate) the null hypothesis in the last column of the table. For each parameter, the pvalue is approximated by the rejection rate of the null hypothesis over the 100 estimations. The results show strong evidence against rejection of null hypothesis in almost all cases.

Figures 5.1, 5.2 and 5.3 show the empirical distributions of the estimates for each parameter. We see that the distribution of the estimates delivered by the EM algorithms are more often unimodal. The parameters of the conditional variance of display bimodal or trimodal densities which might explain why they tend to deviate from their DPG value with higher rejection rate of the null hypothesis.

For the 100 replications, the EM algorithm has inferred the DGP parameters satisfactorily. Based on these results we can confirm that the EM algorithm is able to estimate the path dependent Markov Chain Regime Switching GARCH model. These results also confirm that the EM algorithm performs well.

Parameter	**DGP values**	**Nb of replications**	**Min**	**Max**	**Estimated Mean**	**Percentiles**			**Std errors**	**H0 Rejection Rate**
						25th	50th	75th		
α_1	**0.25**	100	0.21	0.375	**0.26**	0.25	0.26	0.27	0.02436	**0.02**
β_1	**0.0008**	100	-0.0004	0.0015	**0.0007**	0.0005	0.0007	0.0009	0.0004	**0.02**
w_1	**0.5**	100	0.0003	0.514	**0.27**	0.13	0.28	0.44	0.1704	0.23
a_1	**0.35**	100	0.00001	0.99975	**0.32**	0.02	0.16	0.57	0.3379	**0.01**
b_1	**0.40**	100	0.00007	0.995	**0.5**	0.12	0.44	0.73	0.3432	0.21
P_{11}	**0.9739**	100	0.96298	0.9875	**0.9742**	0.972	0.975	0.9765	0.0032	**0.01**
α_2	**-0.07**	100	-0.085935	-0.053	**-0.07**	-0.074	-0.069	-0.065	0.0064	**0.02**
β_2	**-0.0006**	100	-0.000095	-0.0004	**-0.0006**	-0.0007	-0.0006	-0.00057	0.0001	**0.0**
w_2	**0.08**	100	0.00001	0.0833	**0.038**	0.016	0.043	0.048	0.0238	**0.08**
a_2	**0.1**	100	0.00001	0.99968	**0.43**	0.261	0.45	0.546	0.2647	**0.05**
b_2	**0.2**	100	0.00005	0.1685	**0.085**	0.065	0.06	0.126	0.0475	**0.06**
P_{22}	**0.9896**	100	0.986327	0.9921	**0.9894**	0.9885	0.9860	0.9901	0.0012	**0.0**

Table 5.1: **EM Algorithm Simulation Results of the Path Dependent Markov Chain Regime Switching GARCH Model**

Note: Null hypothesis is defined as follow: **H0**: Parameter Estimate = DGP Value; and **H0 Rejection Rate** refers to the probability of rejecting the null hypothesis measured by the number of time the null is rejected divided by the number of replications (100)

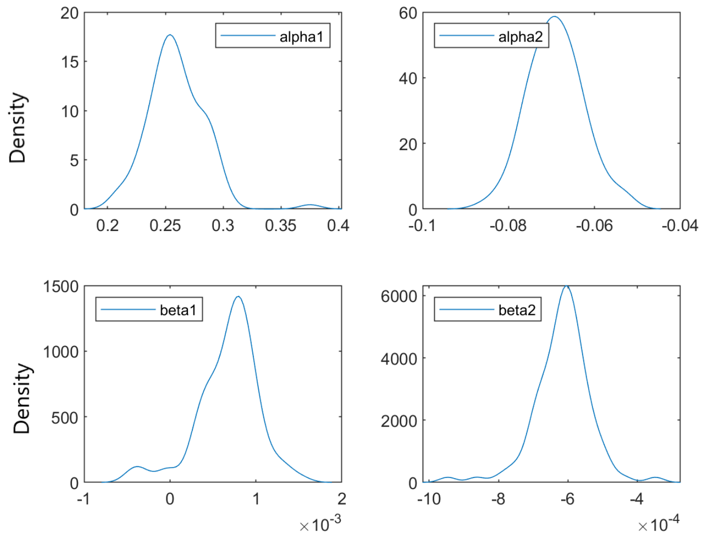


Figure 5.1: **Kernel density of the estimated conditional mean parameters for the Path dependent Regime Switching GARCH Model using the EM algorithm.**

The figure reports the Kernel density distribution of the parameter estimates of the conditional mean of the Path Dependent Regime Switching GARCH Model. We use the EM algorithm to fit the model to the 100 data series simulated with the DGP values from Table 5.1. The left panel reports kernel density distribution of regime 1 parameters and the right panel displays the kernel density distribution of the parameters of regime 2

Note: The conditional mean $\mu_{t,i} = \alpha_i + \beta_i X_{t-1}$; i=1,2.

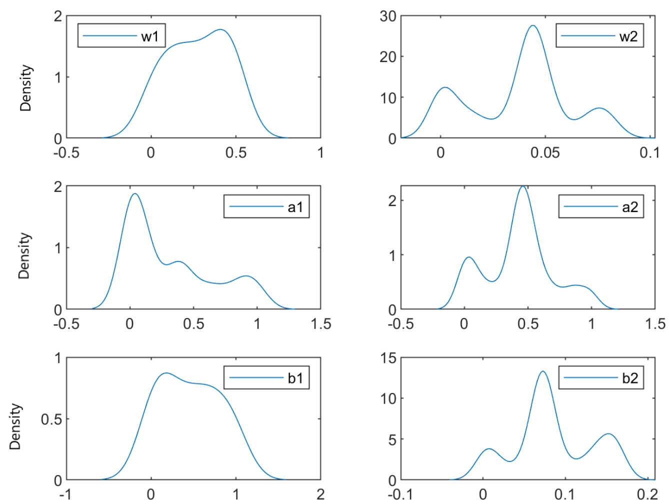


Figure 5.2: **Kernel density of the estimated conditional variances parameters for the Path dependent Regime Switching GARCH Model using the EM algorithm.**

The figure reports the Kernel density distribution of the parameter estimates of the conditional variance of the Path Dependent Regime Switching GARCH Model. We use the EM algorithm to fit the model to the 100 data series simulated with the DGP values from Table 5.1. The left panel reports kernel density distribution of regime 1 parameters and the right panel displays the kernel density distribution of the parameters of regime 2.

Note: The conditional variance $h_{t,i} = w_i + a_i e_{t-1}^2 + b_i h_{t-1,i}$; i=1,2

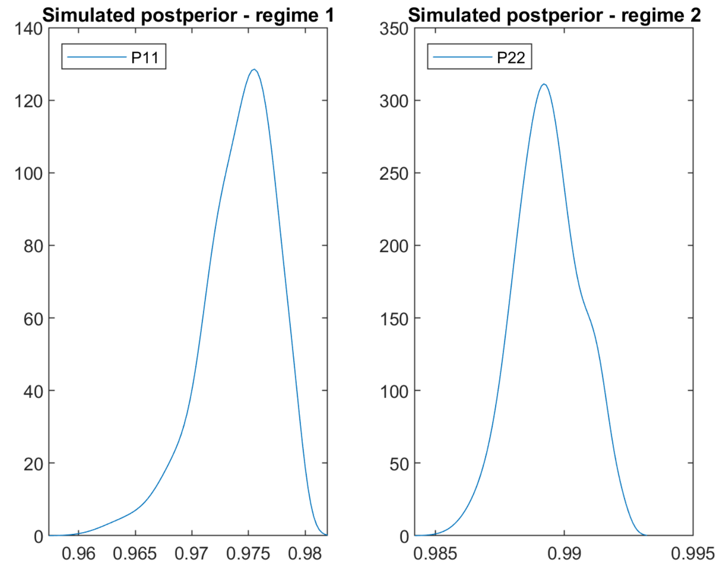


Figure 5.3: **Kernel density of the transition probabilities for the path dependent Regime Switching GARCH Model using the EM algorithm.**

The figure reports the Kernel density distribution of the estimated transition probabilities of the Path Dependent Markov Chain Regime Switching GARCH Model. We use the EM algorithm to fit the model to the 100 data series simulated with the DGP values from Table 5.1. The left panel reports kernel density distribution of the transition probability of regime 1 and the right panel displays the kernel density distribution of the transition probability of regime 2.

Note: The transition probability $P_{i,j}$ is the probability that regime i will be followed by regime j

Chapter 6

Simulation studies for model validation: case of RSG of Gray (1996)

In this section we test whether the RSG model of Gray (Equation (3.1) and (3.2)) would replicate the true parameters when used to estimate a simulated time series with known parameters.

6.1 Simulating of the Markov Chain RSG model of Gray

In this section, we study by simulation the performance of Gray's model. We simulate 100 series of 10000 weekly observations of returns on interest rate and used them to fit the model described in chapter 3 using the Maximum Likelihood Estimation technique outlined in section 4.1. The first subsection lays out the simulation strategy and design. The second subsection discusses the algorithm used to evaluate the objective function and the results of the simulations are presented in the third subsection.

6.1.1 Simulation Strategy and Design

We simulate $M = 100$ samples of size $T = 10000$ from the model (3.1)-(3.2) the same parameters presented in section 5.1 and an interpretation of the parameters can be found in subsection 5.1.1.

$$\begin{aligned}\alpha_1 &= 0.25,\ \beta_1 = 0.0008,\ w_1 = 0.5,\ a_1 = 0.35,\ b_1 = 0.4, \\ \alpha_2 &= -0.07,\ \beta_1 = -0.0006,\ w_2 = 0.08,\ a_2 = 0.1,\ b_2 = 0.2.\end{aligned}$$

The probability of transitions of s_t are:

$$\Pi = \begin{pmatrix} \pi_{11} & \pi_{21} \\ \pi_{12} & \pi_{22} \end{pmatrix} = \begin{pmatrix} 0.9739 & 0.010\,4 \\ 0.026\,1 & 0.9896 \end{pmatrix}.$$

The ergodic probabilities associated with these transition probabilities are:

$$\begin{aligned}\pi_1 &= \frac{1 - 0.9896}{2 - 0.9739 - 0.9896} = 0.285 \\ \pi_2 &= \frac{1 - 0.9739}{2 - 0.9739 - 0.9896} = 0.715\end{aligned}$$

We set the initial values of the variables as follow:

$$\begin{aligned}s_0 &= 1,\, y_0 = 0,\, r_0 = 5 \text{ and} \\ h_0 &= Var\left(y_0 | \text{no information}\right) = 0.005.\end{aligned}$$

where $r_t = r_0 + \sum_{k=1}^{t} y_t$ is the cumulative process.

We assume that $h_{0,1} = h_{0,2} = h_0$ so that y_1 has the same likelihood in both state:

$$f_{1,i} = \frac{1}{\sqrt{2\pi h_{0,i}}} \exp\left(-\frac{y_1^2}{2h_{0,i}}\right) = \frac{1}{\sqrt{2\pi h_0}},\text{ for } i = 1, 2.$$

The posterior probabilities are initialized to the ergodic probabilities:

$$p_{1,1} = \pi_1 = 0.285 \text{ and } p_{1,2} = \pi_2 = 0.715$$

Next, we apply a recursion for $t = 2$ to T. At a given step t, we go through the following sub-steps:

- Step 1: Compute the potential values of the volatility as:

$$\begin{aligned} h_{t,1} &= w_1 + a_1 e_{t-1}^2 + b_1 h_{t-1} \\ h_{t,2} &= w_2 + a_2 e_{t-1}^2 + b_2 h_{t-1} \end{aligned}$$

with the convention that $e_0 = 0$. One of $h_{t,1}$ and $h_{t,2}$ is the realized value of the volatility at time t and both are needed to update h_{t-1}.

- Step 2: Draw a random number U_t from the uniform distribution on $[0,1]$. If $s_{t-1} = 1$, then set $s_t = 1$ if $U_t \leq \pi_{11}$ and $s_t = 2$ otherwise. If $s_{t-1} = 2$, then set $s_t = 2$ if $U_t \leq \pi_{22}$ and $s_t = 1$ otherwise.

- Step 3: Compute $y_t = \mu_t^* + \sqrt{h_t^*}\varepsilon_t$, where ε_t is an independent draw from $N(0,1)$, $\mu_t^* = \alpha_{s_t} + \beta_{s_t} r_{t-1}$, $h_t^* = h_{t,s_t}$. Likewise, compute the cumulative process as $r_t = r_0 + \sum_{k=0}^{t} y_t$ for all t.

- Step 4: Update the posterior probabilities:

$$\begin{aligned} p_{t,1} &= \pi_{11} \frac{f_{t-1,1} p_{t-1,1}}{f_{t-1,1} p_{t-1,1} + f_{t-1,2} p_{t-1,2}} + (1 - \pi_{22}) \frac{f_{t-1,2} p_{t-1,2}}{f_{t-1,1} p_{t-1,1} + f_{t-1,2} p_{t-1,2}} \\ p_{t,2} &= 1 - p_{t,1} \end{aligned}$$

where $f_{t,i} = f\left(r_t | s_t = i, \Omega_{t-1}\right)$ for all t.

- Step: Update h_t and the likelihoods:

$$\begin{aligned} h_t &= p_{t,1}\left(h_{t,1} + \mu_{t,1}^2\right) + p_{t,2}\left(h_{t,2} + \mu_{t,2}^2\right) + \left(p_{t,1}\mu_{t,1} + p_{t,2}\mu_{t,2}\right)^2 \\ f_{t,i} &= \frac{1}{\sqrt{2\pi h_{t,i}}} \exp\left(-\frac{y_t^2}{2h_{t,i}}\right), i = 1, 2 \end{aligned}$$

The next subsection explains how to evaluate the objective functions.

6.1.2 Evaluating the Objective Functions

The sample log-likelihood of the mixture representation $L(\theta)$ given at (4.1) depend on the posterior probabilities $p_{t,1} = \Pr(s_t = i|\Omega_{t-1};\theta)$ and $p_{t,2} = 1 - p_{t,1}$. Indeed, $L(\theta)$ is a sum of logarithms of a weighted average of two likelihood functions.

In order to evaluate the likelihood function $L(\theta)$, we design a MATLAB function called *GRsUnGarch11Gray.m* that takes θ and $\{y_t, r_{t-1}\}_{t=1}^{T}$ as input and returns $p_{t,i}$, $f(y_t|\Omega_{t-1}, s_t = i)$, $h_{t,i}$, h_t, and $L(\theta)$ for $i = 1, 2$ and $t = 1, ..., T$: To begin, we initialize the posterior probabilities at the ergodic probabilities:

$$\begin{aligned} p_{1,1} &\equiv \frac{1-\pi_{22}}{2-\pi_{11}-\pi_{22}}, \\ p_{1,2} &\equiv \frac{1-\pi_{11}}{2-\pi_{11}-\pi_{22}}, \end{aligned}$$

where Ω_0 is an empty set. We have also set the initial conditional variances in each regime to be equal to the sample variance:

$$h_{1,i} \equiv \frac{1}{T-1}\sum_{t=1}^{T}(y_t - \overline{y})^2$$

for $i = 1, 2$, where $\overline{y} = \frac{1}{T-1}\sum_{t=1}^{T} y_t$. Based on these initial values, we compute the first error term and the average conditional variance as:

$$\begin{aligned} e_1 &\equiv y_1 - (p_{1,1}\mu_{1,1} + p_{1,2}\mu_{1,2}), \\ h_1 &\equiv p_{1,1}(h_{1,1} + \mu_{1,1}) + p_{1,2}(h_{1,2} + \mu_{1,2}) - (p_{1,1}\mu_{1,1} + p_{1,2}\mu_{1,2})^2, \end{aligned}$$

where $\mu_{t,i} = \alpha_i + \beta_i r_{t-1}$ for all i and t. We also compute the first conditional likelihoods:

$$f(y_1|\Omega_0, s_1 = i) \equiv \frac{1}{\sqrt{2\pi h_{1,i}}}\exp\left(\frac{-(y_1-\mu_{1,i})^2}{2h_{1,i}}\right).$$

This completes the initialization step. Next, we iteratively compute the following quantities in a loop starting from $t = 2$ and ending at $t = T$:

- Conditional variance per regime:

$$h_{t,i} \equiv w_i + a_i e_{t-1}^2 + b_i h_{t-1},\ \ i = 1, 2.$$

- Likelihood per regime:

$$f\left(y_t | \Omega_0, s_t = i\right) \equiv \frac{1}{\sqrt{2\pi h_{t,i}}} \exp\left(\frac{-\left(y_t - \mu_{t,i}\right)^2}{2h_{t,i}}\right).$$

- Posterior probabilities

$$\begin{aligned}
p_{t,1} &\equiv \pi_{11} \frac{f\left(y_{t-1} | s_{t-1} = 1, \Omega_{t-2}\right) p_{t-1,1}}{\sum_{i=1}^{2} f\left(y_{t-1} | s_{t-1} = i, \Omega_{t-2}\right) p_{t-1,i}} \\
&\quad + \left(1 - \pi_{22}\right) \frac{f\left(y_{t-1} | s_{t-1} = 2, \Omega_{t-2}\right) p_{t-1,2}}{\sum_{i=1}^{2} f\left(y_{t-1} | s_{t-1} = i, \Omega_{t-2}\right) p_{t-1,i}}, \\
p_{t,2} &\equiv 1 - p_{t,1}.
\end{aligned}$$

- Error term and average conditional variance:

$$\begin{aligned}
e_t &\equiv y_t - \left(p_{t,1}\mu_{t,1} + p_{t,2}\mu_{t,2}\right), \\
h_t &\equiv p_{t,1}\left(h_{t,1} + \mu_{t,1}\right) + p_{t,2}\left(h_{t,2} + \mu_{t,2}\right) - \left(p_{t,1}\mu_{t,1} + p_{t,2}\mu_{t,2}\right)^2.
\end{aligned}$$

At the end of this loop, we have all the inputs needed to evaluate the likelihood of the conditional mixture ($L(\theta)$). The code that evaluates the objective function is provided in Appendix.

6.2 Simulation Results

The 100 samples of 10000 observations generated following the simulation strategy presented in section 6.2.1 are used to fit the path independent RSG model of Gray as described in chapter 3. Since the model does not suffer from the usual path dependent problem, it can be simply estimated maximum likelihood.

In Table 6.1, we report the results of the (ML) estimates with some descriptive statistics. The sample mean of the estimated parameters are all close to the DGP value which, fall withing the minimum and maximum bounds of the estimated parameters. However, the null hypothesis that the estimated parameter is equal to the DGP value is rejected for the GARCH coefficients of both equations while the transition probabilities and the coefficients of the conditional mean equations are significantly not different their DGP values.

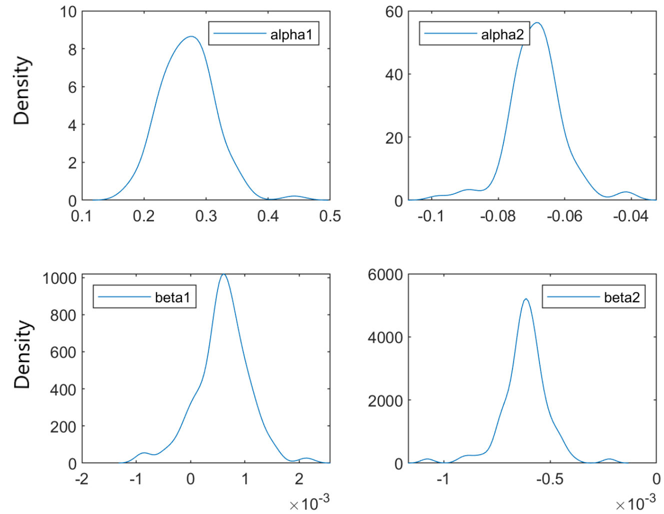
Although, path independent RSG model of Gray does not point identify the DGP values, its solution seems to converge to the same parameter vector. Figures 6.1, 6.2 and 6.3 show the empirical distributions of the estimates for each parameter. We see that the distribution of the estimates delivered by the ML algorithm are all unimodal suggesting that the estimated values of each parameter cluster around a single value.

For the 100 replications, the ML algorithm does not infer the DGP parameters correctly. These results confirm that there is a regime change in the data but the RSG of Gray (1996) is not identified because it fails to identify the true parameters that generate the data. The two regimes are well identified but the path independent Markov Chain Regime Switching Garch model of Gray (1996) is not identified. The identification problem displayed by the model arises from the recombination of the conditional variances as proposed by Gray(1996).

Parameters	DGP values	Nb of replications	Min	Max	Estimated Mean	Percentiles			Std errors	H0 Rejection Rate
						25th	50th	75th		
α_1	**0.25**	100	0.17	0.44	**0.27**	0.24	0.27	0.30	0.043	0.16
β_1	**0.0008**	100	-0.0009	0.0021	**0.0006**	0.0004	0.0006	0.0008	0.001	**0.04**
w_1	**0.5**	100	0.44	0.75	**0.59**	0.549	0.591	0.631	0.056	0.57
a_1	**0.35**	100	0.26	0.44	**0.36**	0.337	0.365	0.388	0.038	0.21
b_1	**0.40**	100	0.20	0.46	**0.32**	0.287	0.323	0.364	0.059	0.49
P_{11}	**0.9739**	100	0.9608	0.9824	**0.9721**	0.9703	0.9721	0.9749	0.004	**0.03**
α_2	**-0.07**	100	-0.098	-0.04	**-0.07**	-0.074	-0.068	-0.065	0.008	**0.01**
β_2	**-0.0006**	100	-0.0011	-0.0002	**-0.0006**	-0.0007	-0.00062	-0.00057	0.0001	**0.03**
w_2	**0.08**	100	0.071	0.100	**0.087**	0.085	0.087	0.090	0.005	0.41
a_2	**0.1**	100	0.0729	0.162	**0.111**	0.097	0.110	0.122	0.017	0.15
b_2	**0.2**	100	0.0504	0.229	**0.135**	0.120	0.134	0.150	0.025	0.70
P_{22}	**0.9896**	100	0.9852	0.9923	**0.9892**	0.9884	0.9893	0.9900	0.001	**0.02**

Table 6.1: **Maximum Likelihood Simulation Results of the Markov Chain Regime Switching GARCH Model**

Note: Null hypothesis is defined as follow: **H0**: Parameter Estimate = DGP Value; and **H0 Rejection Rate** refers to the probability of rejecting the null hypothesis measured by the number of time the null is rejected divided by the number of replications (100)



...

Figure 6.1: **Kernel density of estimated conditional mean parameters of the Regime Switching GARCH Model of Gray(1996) using Maximum Likelihood Estimation.**

The figure reports the Kernel density distribution of the parameter estimates of the conditional mean of the Regime Switching GARCH Model of Gray(1996). We use the ML algorithm to fit the model to the 100 data series simulated with the DGP values from Table 6.1. The left panel reports kernel density distribution of regime 1 parameters and the right panel displays the kernel density distribution of the parameters of regime 2.

Note: The conditional mean $\mu_{t,i} = \alpha_i + \beta_i X_{t-1}$; i=1,2

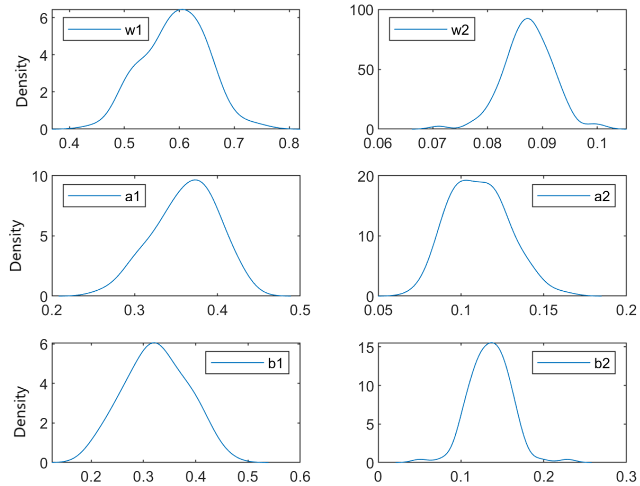


Figure 6.2: **Kernel density of estimated conditional variances parameters of the Regime Switching GARCH Model of Gray(1996) using Maximum Likelihood Estimation.**

The figure reports the Kernel density distribution of the parameter estimates of the conditional variance of the Regime Switching GARCH Model of Gray(1996). We use the ML algorithm to fit the model to the 100 data series simulated with the DGP values from Table 6.1. The left panel reports kernel density distribution of regime 1 parameters and the right panel displays the kernel density distribution of the parameters of regime 2.

Note: The conditional variance $h_{t,i} = w_i + a_i e_{t-1}^2 + b_i h_{t-1}$; i=1,2

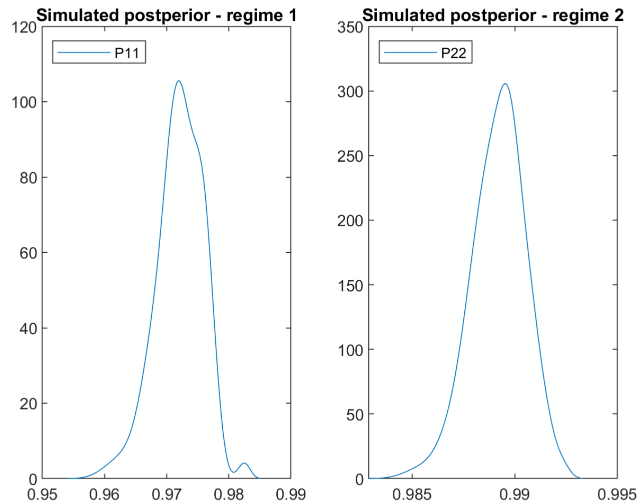


Figure 6.3: **Kernel density of the estimated transition probabilities of the Regime Switching GARCH Model of Gray(1996) using Maximum Likelihood Estimation.**

The figure reports the Kernel density distribution of the estimated transition probabilities of the Markov Chain Regime Switching GARCH Model of Gray(1996). We use the ML algorithm to fit the model to the 100 data series simulated with the DGP values from Table 6.1. The left panel reports kernel density distribution of the transition probability of regime 1 and the right panel displays the kernel density distribution of the transition probability of regime 2.

Note: The transition probability $P_{i,j}$ is the probability that regime i will be followed by regime j

Chapter 7

Empirical Application: Fitting the Regime Switching GARCH

7.1 The Data

For this empirical application, we use a time series of weekly one-month Treasury bill rates observed from January 1970 to April 1994, ($T = 1267$ observations). The same data have been used in the empirical section of Gray (1996). The interest rate are expressed in annualized percentage, meaning that a 5% annualized interest rate is recorded as $r_t = 5$. Subsequently, we let r_t denotes the interest rate process. Table 2 shows descriptive statistics for the interest rate (r_t) and its first difference (Δr_t). The average interest rate on the period covered by the sample is 7.068%, with a standard deviation of 2.802%. The maximum rate observed during this period is 17.01%. The interest rate process is positively skewed and slightly fat tailed. The first difference process (Δr_t) has a relatively large standard deviation and negative skewness, and its kurtosis is several times as high as that of r_t.

Descriptive Statistics	r_t	Δr_t
Mean	7.0682	-0.0034
Standard Dev.	2.8021	0.3496
Min	2.7300	-4.2200
Max	17.0100	2.2200
Skewness	1.0089	-1.5116
Kurtosis	4.1034	28.3817

Table 7.1: **Descriptive Statistics of weekly one-month Treasury bill rates from January 1970 to April 1994**

Figure 5 shows the time series plots of r_t (left) and $y_t = \Delta r_t$ (right). These graphs are the same as those in Figure 3 of Gray (1996). It is seen on the panel at the right that Δr_t exhibits some jumps. We do not address the presence of these jumps in the data because our objective is to assess the performance of the same RSG model as in Gray (1996).

Finally, Figure 6 shows the trajectory of the sample variance of y_t over time based on a 24 weeks rolling windows. It is seen that that high periods of volatility are followed by low periods of volatility, much like in our previous simulation. This is not surprising since these simulations are done using parameter values that are estimated by Gray (1996) from the current data set.

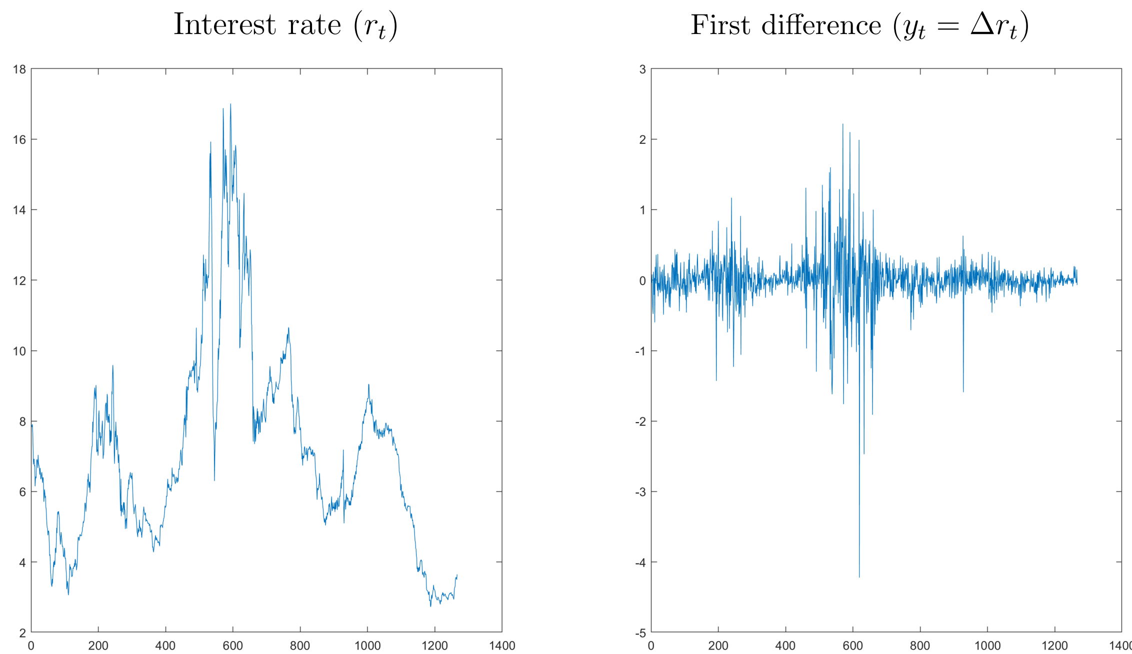


Figure 7.1: **Plot of weekly one-month Treasury bill rates r_t from January 1970 to April 1994 the data**

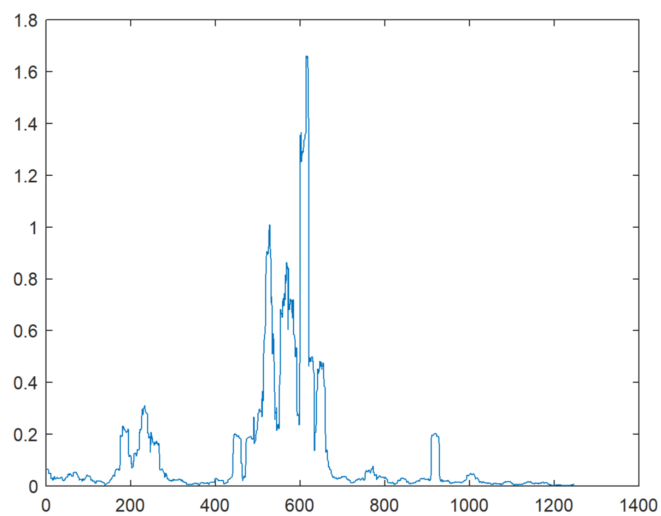


Figure 7.2: **Sample variance of $y_t = \Delta r_t$ based on 24 weeks rolling windows**

7.2 Estimation Results of the RSG of Gray using the ML algorithm

As a first step of the assessment of the RSG model proposed by Gray, we attempt to replicate the results of the Gray (1996) paper except for the Generalized RSG models. We fit the RSG model of Gray (1996) to the data using the Maximum likelihood algorithm. Next, we fit the path dependent Regime Switching GARCH (henceforth PD-RSG) model to the same using the EM algorithm. The estimation of each model is done in three steps.

1. First, we draw independently a sample of 100 000 initial set of values from the space of parameters where each parameter belongs to its domain of definition.
2. Second, we evaluate the objective function 100 000 times with each set of parameters and select the value of the objective function that minimizes the negative log-likelihood.
3. Finally, the parameter obtained in step 2 is used as starting to the optimization algorithm. This last step is repeated until the optimization algorithm reaches a satisfactory[1] local minimum using the latest estimated parameter vector as the starting point for the following optimization.

We start by replicating four models from Gray's paper starting from the simple model with constant variance to the full Regime Switching GARCH models. Next we move to the simulation studies.

The results of the replication of Gray's results are summarized in Table 7.2 and 7.3. Tables 7.2 reports results of two constant variance models while Tables 7.3 reports the results of two GARCH models. The results show that overall, our programs replicate Gray's outcomes.

Table 7.2 reports the results of the estimation of a single regime and a markov

[1]Exiflag =1 or Exitflag =2 and the gradian close (or sufficiently close) to zero and all other optimizations constraints are respected.

chain regime switching constant variance model. In the first four column are the results of the estimations of a single regime model with a constant variance (Equations (2.1) and (2.2) with $a_i = 0$ for $i = 1..k$). The last four column of the table display the results of the estimation of a regime switching with constant variance (Equation (3.1) and (3.2) with $a_{s_t} = b_{s_t} = 0$). For both single regime and regime switching cases, our programs converge to a local minimum with log-likelihood slightly higher and the parameters estimates close to Gray's results.

The Table 7.3 displays the results of two models (sigle regime and regime switching) with GARCH error terms. from the estimation of the RSG models (Equation (3.1) and (3.2)). In the case of the single RSG model or the simple GARCH model, our estimates are close to Gray's and although our log-likelihood is slightly smaller, the first order optimality conditions were fully met. Our programs perform better in the case of the full RSG model as the maximum log-likelihood from our optimization process is higher than Gray's and the estimated parameters seem more accurate. These findings suggest that our codes perform well. In the next section we move to assess the validity of the RSG models of Gray (1996) through simulations studies.

	Single regime with constant variance				**Regime switching with constant variance**			
	Own calculation		Gray 1996		Own calculation		Gray 1996	
Parameters	Estimate	T Statistics	Estimate	T Statistics	Estimate	T Statistics	Estimate	T Statistics
α_1	0.0475	1.7839	0.0475	1.1757	0.1673	0.0576	0.1687	1.2102
α_2					0.0026	0.0143	-0.0057	-0.3132
β_1	-0.0072	-2.0567*	-0.0072	-1.0746	-0.0189	-0.0765	-0.0190	-1.2752
β_2					0.0001	0.0031	0.0015	0.4688
w_1	0.1218	0.0054	0.3489	13.7362*	0.4489	0.197	0.6716	8.5762*
w_2					0.0223	0.363	0.1496	19.6842*
P					0.9686	9.0324*	0.9698	53.4807*
Q					0.9907	11.0626*	0.9905	202.1429*
Log-likelihood	-463.276		-463.143		111.1344		111.1109	
First order Optimality	4.36E-04				8.54E-05			
Exitflag	1				1			

* Significant at 5% or less
Index 1-Regime1
Index 2-Regime2

Table 7.2: **Results of the estimation of a single regime and a markov chain regime switching constant variance model.**

The model is fit with a data sample of 1267 observations of weekly one-month Treasury bill yields in annualized percentage terms from January 1970 to April 1994. The first two columns report the results of our estimations of a single-regime GARCH model while columns 3 and 4 display the results of the same model from Gray(1996), Table 2 (column 1 and 2). Columns 5 and 6 display the results of our estimations of a markov chain regime switching GARCH model while the last two columns report the results of the same model from Gray(1996) Table 2 (column 3 and 4).

	Single regime with Garch error term				Regime switching with Garch error term			
	Own calculation		Gray 1996		Own calculation		Gray 1996	
Parameters	Estimate	T Statistics	Estimate	T Statistics	Estimate	T Statistics	Estimate	T Statistics
α_1	0.0054	0.4011	0.0059	0.5017	0.1419	1.3301	0.1407	1.1616
α_2					0.0030	0.2044	-0.0011	-0.0762
β_1	-0.0012	-0.5414	-0.0013	-0.516	-0.0142	-1.426	-0.0141	-1.0807
β_2					-0.0001	-0.0411	0.0006	0.2366
w_1	0.0007	3.0553*	0.0004	1.1356	0.1853	3.2398*	0.1870	2.2278*
w_2					0.0099	5.5473*	0.4609	1.9885*
a_1	0.1632	7.9044*	0.2095	2.5506*	0.4627	3.2419*	0.1977	1.2378
a_2					0.1654	3.3575*	0.0099	3.861*
b_1	0.8325	39.7376*	0.8208	13.4603*	0.2001	1.5653	0.1655	2.7556*
b_2					0.2692	4.0375*	0.2685	2.815*
P					0.9742	75.333*	0.9739	24.1357*
Q					0.9897	237.5014*	0.9896	86.8336*
Log-likelihood	124.5016		129.7304		151.3096		151.2288	
output.firstorderopt	7.63E-06				6.65E-05			
Exitflag	2				1			

* Significant at 5% or less
Index 1-Regime1
Index 2-Regime2

Table 7.3: **Results of the estimation of a single regime and a markov chain regime switching GARCH model**

The model is fit with a data sample of 1267 observations of weekly one-month Treasury bill yields in annualized percentage terms from January 1970 to April 1994. The first two columns report the results of our estimations of a single-regime GARCH model while columns 3 and 4 display the results of the same model from Gray(1996), Table 3 (column 1 and 2). Columns 5 and 6 display the results of our estimations of a markov chain regime switching GARCH model while the last two columns report the results of the same model from Gray(1996) Table 3(column 3 and 4).

7.3 Estimation Results of the PD-RSG model using the EM algorithm

In this section, we fit the path dependent Regime Switching GARCH (henceforth PD-RSG) model to the same using the EM algorithm. Considering the presentation in chapter 2 and 3, the RSG model of Gray (1996) resulted from the a modification of the PD-RSG. The models are different and are expected to yield different parameters estimates. Recall that by construction, the EM algorithm does not necessarily converge to the global maximum. To mitigate this risk, we estimation the model in three steps as follow.

1. First, we draw independently a sample of 100 000 initial set of values from the space of parameters where each parameter belongs to its domain of definition.
2. Second, we evaluate the objective function 100 000 times with each set of parameters and select the value of the objective function that minimizes the negative log-likelihood.
3. Finally, the parameter obtained in step 2 is used as starting to the optimization algorithm. This last step is repeated until the optimization algorithm reaches a satisfactory[1] local minimum using the latest estimated parameter vector to evaluate the posterior probabilities of the **E-Step** for the following optimization.

Alternatively, we use the parameter estimates of the Gray's model as an informed guess of the starting point of the algorithm. This approach leads to better results as the convergence criteria of the EM algorithm is smaller and the log-likelihood is higher.

Table 7.5 display the results of the estimations to the path dependent RSG model with the EM algorithm. The results are close to Gray's estimates. However, most of the parameter estimates are not statistically different from zero.

[1]The convergence criteria of the EM algorithm is defined as the distance between two consecutive vectors of parameter estimates to be less that 0.001.

	Regime switching with constant probabilities			
	Own Calculation PD-RSG EM		Gray (1996) MLE	
Parameters	Estimate	T Statistics	Estimate	T Statistics
α_1	0.1991	5.7887*	0.1407	1.1616
α_2	-0.0868	-3.6408*	-0.0011	-0.0762
β_1	0.0036	0.1960	-0.0141	-1.0807
β_2	-0.1174	-0.1186	0.0006	0.2366
w_1	0.1868	0.2561	0.1870	2.2278*
w_2	0.0113	0.0989	0.4609	1.9885*
a_1	0.4606	0.4449	0.1977	1.2378
a_2	0.1662	0.2109	0.0099	3.861*
b_1	0.1975	0.1929	0.1655	2.7556*
b_2	0.2682	0.2899	0.2685	2.815*
P	0.9886	1.5913	0.9739	24.1357*
Q	0.9729	3.7000*	0.9896	86.8336*

* Significant at 5% or less
Index 1-Regime1
Index 2-Regime2

Table 7.4: **Results of the Estimation of the Path Dependent Markov Chain RSG Model using the EM algorithm.**

The model is fit with 1267 observations of weekly one-month Treasury bill yields in annualized percentage terms from January 1970 to April 1994. The first two columns report the results of our estimations while the last two display the results from Gray(1996), Table 3.

Chapter 8

Conclusion

There is a growing interest in using Markov Regime Switching GARCH (RSG) models to analyze and forecast volatilities financial markets. However Maximum Likelihood Estimation (MLE) of the path dependent RSG models is infeasible in practice. Gray(1996) developed a generalized MSG model that solves the path dependence issue and rendering the MLE possible. Following Gray(1996), several versions of the Regime Switching Garch models that can be estimated by a maximum likelihood method have been proposed; see (Duerke (1997), Klaassen(2002), Haas et al.(2004)). Rather than modifying the original path dependent RSG, other researched have proposed a Bayesian Markov Chain Monte Carlo (MCMC) algorithm that overcomes the path dependence; see Das and Yoo (2004), Bauwens, Preminger and Rombouts (2010), Henneke, Rachev and Fabozzi (2011).

This thesis contributes to this growing literature in two ways.

First, this paper studies the empirical properties of the Regime Switching GARCH (RSG) of the Gray (1996) model. This model is appealing because it does not exhibit the path dependent that is typical to most RSG models. We investigate by simulations the ability of the Maximum Likelihood method (based on mixture representation) to identify the parameters of this model. To begin, we validate our programs using common random number simulations to assess the ability of ours programs to reproduce the parameters used to simulate the data. The second validation technique consisting in comparing the results of our estimation with the outputs of popular packages in MatLab or Stata.

Subsequently, with simulations studies, we assess the empirical properties of the Markov Regime Switching GARCH (RSG) model of Gray (1996). We simulate 100 data series with the RSG model proposed by Gray (1996). We then fit the model with the simulated data using the Maximum Likelihood Estimation (MLE). Our results show that the RSG model of Gray (1996) does not infer the DGP parameters correctly. These results confirm that model fails to identify the true parameters that generate the data. We conclude that the path independent Markov Chain Regime Switching GARCH model of Gray/(1996) is not identified. The failure of the RSG model of Gray (1996) to infer true parameters from simulated data can only be linked to its theoretical properties particularly, the approximation of the true log likelihood by collapsing the recursive process of the GARCH. Our findings echoes previous research by Haas, Mittnik and Paollella (2004) who also raised concerns about the consequences of recombination of the conditional variances.

Second, we also investigate by simulations the ability of the EM algorithm to identify the DGP parameters of the path dependent Regime Switching GARCH (PD-RSG) model. We shows that the common path dependence problem of RSG models can be solved with the Expectation-Maximization Algorithm (EM). We simulate 100 data series with the path dependent RSG model with known parameters. We then use the EM algorithm to fit the model with the simulated data. Our results confirm that the EM algorithm infers the Data Generating Process (DGP) parameters satisfactorily and we conclude that the algorithm is able to estimate the path dependent Markov Chain Regime Switching GARCH models.

The merite of this thesis is to be the first analytical to show empirically (i) the limits of the Markov Regime Switching GARCH (RSG) model of Gray (1996) and (ii) to use the EM algorithm to fit the path dependent RSG model to data. We show that the EM is an alternative approach and perhaps an easier way of fitting the path dependent RSG models. Against these backdrops, it would not be advisable to perform financial markets volatilities analyzing based on a RSG model of the family of the RSG models of Gray (1996).

AS next steps, next researches could compare the performance of the EM algorithm and MCMC for the estimation of path dependent the RSG models. Also, empirical properties of modified versions of the original path dependent RSG should be assessed. Finally, an extension of the use of the EM algorithm to estimate bivariate RSG models would be useful especially to derive optimal hedging ratios.

References

Andersen, T.G. and Bollerslev, T. (1998), "Answering the Skeptics: Yes, Standard Volatility Models do Provide Accurate Forecasts," *International Economic Review*, 39, 885-905.

Bauwens, Dufays and Rombouts (2014), "Marginal Likelihood for Markov Switching and Change-Point GARCH Models," Journal of Econometrics, Vol 178, Part 3, January 2014, Pages 508–522.

Bauwens, Preminger and Rombouts (2010), "Theory and inference for a Markov switching GARCH model," *The Econometrics Journal*, Vol 13, Issue 2, pages 218–244, July 2010

Bollerslev, T. "Generalized Autoregressive Conditional Heteroskedasticity," *Journal of Econometrics*, vol. 31, 1986, p. 307–327

Cai, J. (1994), "A Markov Model of Switching-Regime ARCH," *Journal of Business Economic Statistics*, 12, p. 309-316

Das, D., Yoo, B. H. (2004) "A Bayesian MCMC Algorithm for Markov Switching GARCH models, " *Econometric Society 2004*, North American Summer Meetings 179, Econometric Society, 2004.

Diebold, F. (1986), "Comment on Modeling the Persistence of Conditional Variances," *Econometric Reviews.* 5, 51â56.

Dempster, A. P., Laird, N.M., and Rubin, D.B. (1977), "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, Series B (Methodological), Vol. 39, No. 1. (1977), pp. 1-38.

Dueker, M. (1997), "Markov Switching in GARCH Processes in Mean Reverting Stock Market Volatility, " *Journal of Business and Economics Statistics*, 15, 26â34.

Engle, R. (1982), "Autoregressive Conditional Heteroskedasticity With Estimates of the Variance of U.K. Inflation," *Econometrica* 50 (1982): 987-1008.

Gray, S.F. (1996), "Modeling the conditional distribution of interest rates as a regime-switching process," *Journal of Financial Economics* 42, Issue 1, September 1996, Pages 27–62.

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

Hamilton, J. D., and Susmel, R. (1994), "Autoregressive Conditional Heteroskedasticity and Changes in Regime, " *Journal of Econometrics*, 64, 307-333

Kim, C. J. (1994), "Dynamic Linear Models With Markov Switching," *Journal of Econometrics*, 64, 1-22

Klaassen, F. (2002), "Improving GARCH Volatility Forecasts with Regime-Switching GARCH, " *Universiteit Tubingen, Working Paper.*

Lamoureux, C., and W. Lastrapes (1990), "Heteroskedasticity in Stock Return Data: Volume versus GARCH Effects, " *Journal of Finance*, 45, 221â229.

Navidi, W. (1997), "A graphical illustration of the EM Algorithm," *The American Statistician*, Vol 51, No. 1, pp. 29-31.

Appendix A

MATLAB CODES

Copy and paste these codes into a MATLAB ”m-file”. Then, use CTRL+A to select all and CTRL+I to indent.

A. The function ”Data_Sim_RSGarch11.m”

This function simulates the 100 series of 10000 observations for the path dependent Markov Chain Regime Switching GARCH(1,1) model.

```
% Data_Sim_RSGarch11
%% Parameters
clc
clear

% Sample size
T = 10000;

% State variable
S = zeros(T+1,1);
S(1) = 1; % assumes "S0=1"

```

```
% Interest rate
r = zeros(T+1,1);
r(1) = 0.05; % short rate is initialized at 5%

% Increments of interest rate
y = zeros(T,1);

% Prior transition probabilities
p11 = 0.9739;
p22 = 0.9896;


alpha1 = 0.25;
beta1 = 0.0008;
w1 = 0.50;
a1 = 0.35;
b1 = 0.40;
% a1 = 0;
% b1 = 0;

alpha2 = -0.07;
beta2 = -0.0006;
w2 = 0.08;
a2 = 0.10;
b2 = 0.20;
% a2 = 0;
% b2 = 0;
%%
% Conditional variance per regime
hi = zeros(T,1);

h1 = zeros(T,1);
h2 = zeros(T,1);

h1(1)= w1/(1-(a1+b1));
h2(2)= w2/(1-(a2+b2));

% Conditional means

mu1 = zeros(T,1);
```

```
mu2 = zeros(T,1);
mui= zeros(T,1);

% Likelihoods
g1 = zeros(T+1,1);
g2 = zeros(T+1,1);

% Likelihood of r0 for each state
g1(1) = 1/sqrt(2*pi*h1(1));
g2(1) = 1/sqrt(2*pi*h2(1));

% Posterior probabilities
p1 = zeros(T+1,1);
p2 = zeros(T+1,1);

% qq1 = zeros(T+1,1);
% Unconditional regime probabilities
% p1(1) = 0.285;
p1(1)=(1-p22)/(2-p11-p22);
p2(1) = 1 - p1(1);

% e = 0;
e = zeros(T,1);
e(1)= randn(1);

%% Simulation
NbSim = 100;
SimY_full = zeros(T,NbSim);
SimX_full = zeros(T,NbSim);

for i =1:1:NbSim

for t=2:T+1

% Conditional mean per state
mu1(t-1) = alpha1 + beta1*r(t-1);
mu2(t-1) = alpha2 + beta2*r(t-1);

% Conditional variance per state
h1(t) = w1 + a1*e(t-1)^2 + b1*h1(t-1);
```

```
h2(t) = w2 + a2*e(t-1)^2 + b2*h2(t-1);

% Current state and increment of interest rate: y(t)
u = rand;
v1 = randn;
v2 = randn;
if S(t-1) == 1
if u < p11
S(t) = 1;
hi(t) = h1(t);
mui(t-1) = mu1(t-1);
y(t) = mu1(t-1) + sqrt(hi(t))*v1;
%              y(t) = normpdf(0,mu1(t-1),sqrt(hi(t-1)));
else
S(t) = 2;
hi(t-1) = h2(t);
mui(t-1) = mu2(t-1);
y(t) = mu2(t-1) + sqrt(hi(t))*v2;
%              y(t) = normpdf(0,mu2(t-1),sqrt(hi(t-1)));
end
elseif S(t-1) == 2
if u < p22
S(t) = 2;
hi(t) = h2(t);
mui(t-1) = mu2(t-1);
y(t) = mu2(t-1) + sqrt(hi(t))*v2;
%              y(t) = normpdf(0,mu2(t-1),sqrt(hi(t-1)));
else
S(t) = 1;
hi(t) = h1(t);
mui(t-1) = mu1(t-1);
y(t) = mu1(t-1) + sqrt(hi(t))*v1;
%              y(t) = normpdf(0,mu1(t-1),sqrt(hi(t-1)));
end
end

% Interest rate
r(t) = r(t-1) + y(t);

% likelihood
```

```

g1(t-1) ...
    =(1/sqrt(2*pi*h1(t-1)))*exp((-1/(2*h1(t-1)))*(y(t)-mu1(t-1))^2);
g2(t-1) ...
    =(1/sqrt(2*pi*h2(t-1)))*exp((-1/(2*h2(t-1)))*(y(t)-mu2(t-1))^2);

% Regime posterior probabilities
p1(t) = (1-p22)*g2(t-1)*(1-p1(t-1))/(g1(t-1)*p1(t-1) + ...
    g2(t-1)*(1-p1(t-1)))...
+ p11*g1(t-1)*p1(t-1)/(g1(t-1)*p1(t-1) + g2(t-1)*(1-p1(t-1)));

%    p1(t) = ...
    (1-p22)+((p11+p22-1)*(g1(t)*p1(t-1))/(g1(t)*p1(t-1)+g2(t)*p2(t-1)));
p2(t) = 1 - p1(t);


% error
e(t) = y(t) - mui(t-1);


end
SimY_full(:,i)= y(2:end);
SimX_full(:,i)= r(1:end-1);
end
```

B. The function "ProbRsGarchPD.m"

This function evaluates the posterior probabilities and conditional likelihoods. The outputs of the function are used in the EM approaches.

```

function [Pold,Giold] = ProbRsGarchPD(parameter,Y,X1)
% % % E-Step:
% - Imputs: parameter,Y,X1
% parameter: at t+1, parameter = estimated parameter at ...
    M-step at time t
% - Pold,Giold
% Pold : Posterior Probabilities per regime given parameter
% Giold : The likelihood function per regime given parameter

% the parameters
k1 = size(X1,2);
N=2;
np=6;
param =zeros(np,N);

param(1:end,1)=parameter(1:np);
param(1:end,2)=parameter(np+1:end);

alpha       = param(1,:);
beta        = param(2:k1+1,:);
w           = param(k1+2,:);
a           = param(k1+3,:);
b           = param(k1+4,:);
P = param(end,1);
Q = param(end,2);

%---------------------------------------------------------------------------
T = length(Y);
%---------------------------------------------------------------------------
%% Initial Values
% Initial values for the regime probabilities epsilun
% ...
```

```
    --------------------------------------------------------------------------
epsilun = zeros(T+1,N); % It maybe better to take this as a ...
    N*1 vector
epsilun(1,:)=  [(1-Q)/(2-P-Q) , (1-(1-Q)/(2-P-Q))]; % the ...
    steady state probabilities
%-------------------------------------------------------------------------
% The conditional mean mu by regime
%-------------------------------------------------------------------------
mu = zeros(T,N);  % It maybe better to take this as a N*1 ...
    vector
% I compute12d the mu for each obs by regime
for i=1:N
mu(:,i) = ones(T,1)*alpha(i)+X1*beta(:,i);
end

%-------------------------------------------------------------------------
%Computing the conditional variances by regime "hit" for ...
    each period
%-------------------------------------------------------------------------
% Conditionnal variances
coefH = [w' a' b'];
Hi = zeros(T,N);
%Hi(1,:)= w./(ones(1,N)-a-b);
Hi(1,:)=var(Y);

% Error term

Error =  zeros(T,1);
Error(1) = Y(1)-(mu(1,:)*epsilun(1,:)');

% The likelihood function by regime
Gi = zeros(T,N);
% Gi(1,:)=(1./sqrt(2*pi*Hi(1,:))); % when Error(1)=0
Gi(1,:)=(1./sqrt(2*pi*Hi(1,:))).*exp((-1./(2*Hi(1,:))).*(Y(1)-mu(1,:)).^2);

for t=2:T

% Hi conditional variances
Hi(t,:) =[1  Error(t-1)*Error(t-1)  Hi(t-1)]*coefH' ;

```

```
% The loglikelihood function
Gi(t,:)=(1./sqrt(2*pi*Hi(t,:))).*exp((-1./(2*Hi(t,:))).*(Y(t)-mu(t,:)).^2);
%     ...
    Gi(t-1,:)=(1./sqrt(2*pi*Hi(t-1,:))).*exp((-1./(2*Hi(t-1,:))).*(Y(t-1)-mu(t-1,
% the next period regime probabilities
%    epsilun(t,1)= ...
    (1-Q)+((P+Q-1)*(Gi(t,1)*epsilun(t-1,1))/(Gi(t,:)*epsilun(t-1,:)'));

epsilun(t,1)= ...
    (1-Q)*((Gi(t-1,2)*epsilun(t-1,2))/(Gi(t-1,:)*epsilun(t-1,:)')) ...
    ...
+ P*((Gi(t-1,1)*epsilun(t-1,1))/(Gi(t-1,:)*epsilun(t-1,:)'));
epsilun(t,2)= 1-epsilun(t,1);
% error term
Error(t)= Y(t)-(mu(t,:)*epsilun(t,:)');
end
Pold=epsilun(1:T,:);
Giold=Gi(1:T,:);
```

C. The function ”EM$_{R}sGarchPD.m$”

This function evaluates the auxiliary function for the EM algorithm .

```
function [EM_MaxRsGarchPD] = EM_RsGarchPD(parameter,Y,WW)
%% M-Step imputs
X1= WW(:,1); % is r(t-1) of the conditional mean equation
Pold = WW(:,2:3); % Posterior Probabilities from the E-Step
%Giold = WW(:,4:5); % The likelihood function from the E-Step
%%
% the parameters
k1 = size(X1,2);
N=2; % nb of regimes
np=6; % nb of parameters per regime
param =zeros(np,N);

param(1:end,1)=parameter(1:np);
param(1:end,2)=parameter(np+1:end);

alpha      = param(1,:);
beta       = param(2:k1+1,:);
w          = param(k1+2,:);
a          = param(k1+3,:);
b          = param(k1+4,:);
P = param(end,1);
Q = param(end,2);
%% Current Posterior Probabilities
% Initial Values
%--------------------------------------------------------------------------
T = length(Y);
%--------------------------------------------------------------------------
% Initial values for the regime probabilities epsilun
% ...
    --------------------------------------------------------------------------
epsilun = zeros(T+1,N); % It maybe better to take this as a ...
    N*1 vector
epsilun(1,:)=  [(1-Q)/(2-P-Q) , (1-(1-Q)/(2-P-Q))]; % the ...
    steady state probabilities
```

```
%------------------------------------------------------------------------------------------
% The conditional mean mu by regime
%------------------------------------------------------------------------------------------
mu = zeros(T,N);
for i=1:N
mu(:,i) = ones(T,1)*alpha(i)+X1*beta(:,i);
end

%------------------------------------------------------------------------------------------
%Computing the conditional variances by regime "hit" for ... each period
%------------------------------------------------------------------------------------------
% Conditionnal variances
coefH = [w' a' b'];
Hi = zeros(T,N);
%Hi(1,:)= (w + a)./(ones(1,N)-b);
Hi(1,:)=var(Y);

% Error term

Error =  zeros(T,1);
Error(1) = Y(1)-(mu(1,:)*epsilun(1,:)');

% The likelihood function by regime
Gi = zeros(T,N);
%Gi(1,:)=(1./sqrt(2*pi*Hi(1,:)));
Gi(1,:)=(1./sqrt(2*pi*Hi(1,:))).*exp((-1./(2*Hi(1,:))).*(Error(1)).^2);


for t=2:T

% Hi conditional variances
Hi(t,:) =[1  Error(t-1)*Error(t-1)  Hi(t-1)]*coefH' ;

% The loglikelihood function
%   ... Gi(t,:)=(1./sqrt(2*pi*Hi(t-1,:))).*exp((-1./(2*Hi(t-1,:))).*(Y(t)-mu(t,:)).^2
Gi(t,:)=(1./sqrt(2*pi*Hi(t,:))).*exp((-1./(2*Hi(t,:))).*(Y(t)-mu(t,:)).^2);
% the next period regime probabilities
%    epsilun(t,1)= ...
```

```
    (1-Q)+((P+Q-1)*(Gi(t,1)*epsilun(t-1,1))/(Gi(t,:)*epsilun(t-1,:)'));
epsilun(t,1)= ...
    (1-Q)*((Gi(t-1,2)*epsilun(t-1,2))/(Gi(t-1,:)*epsilun(t-1,:)')) ...
    ...
+ P*((Gi(t-1,1)*epsilun(t-1,1))/(Gi(t-1,:)*epsilun(t-1,:)'));
epsilun(t,2)= 1-epsilun(t,1);
% error term
Error(t)= Y(t)-(mu(t,:)*epsilun(t,:)');
end

Pnew=epsilun(1:T,:); % Current Posterior Probabilities

Qn = Pold.*log(((Pnew.*Gi)./Pold));
EM_MaxRsGarchPD = -sum(Qn*ones(N,1));

%% Auxiliary objective function
% RsGarchPD2 = log(Pnew.*Gi)*ones(N,1);
% EM_MaxRsGarchPD= -sum((Pold.*RsGarchPD2)*ones(N,1));

%%
EM_MaxRsGarchPD(isnan(EM_MaxRsGarchPD))=1e6;
EM_MaxRsGarchPD(isinf(EM_MaxRsGarchPD))=1e6;
```

C. The function "Simulation$_{Gray_{s}hort_{r}ate.m}$"

This function simulate data using RSG omodel of Gray (1996)

```
%% Parameters
clc
clear

% Sample size
T = 10000;

% State variable
S = zeros(T+1,1);
S(1) = 1; % assumes "S0=1"

% Interest rate
r = zeros(T+1,1);
r(1) = 7.0682; % short rate is initialized at 5%

% Increments of interest rate
y = zeros(T,1);

% Prior transition probabilities
p11 = 0.9739;
p22 = 0.9896;

% Posterior probabilities
p1 = zeros(T+1,1);
p2 = zeros(T+1,1);

p1(1)=(1-p22)/(2-p11-p22);
p2(1) = 1 - p1(1);

%
alpha1 = 0.25;
beta1 = 0.0008;
w1 = 0.50;
a1 = 0.35;
```

```
b1 = 0.40;
% a1 = 0;
% b1 = 0;

alpha2 = -0.07;
beta2 = -0.0006;
w2 = 0.08;
a2 = 0.10;
b2 = 0.20;
% a2 = 0;
% b2 = 0;

% variances
h0 = 0.005; % first observation is Var(r0|no information)

% Error term
e = 0;

% Conditional variance per regime
h1 = zeros(T,1);
h2 = zeros(T,1);

h1(1) = w1 + a1*e^2 + b1*h0;
h2(1) = w2 + a2*e^2 + b2*h0;

% Conditional variances (regime independent)
h = zeros(T+1,1);
h(1) = h0;

% Likelihoods
g1 = zeros(T,1);
g2 = zeros(T,1);



%% Simulation
NbSim = 100;
SimY_full = zeros(T,NbSim);
SimX_full = zeros(T,NbSim);

```

```
for i =1:1:NbSim

for t=2:T+1

% Conditional mean per state
mu1 = alpha1 + beta1*r(t-1);
mu2 = alpha2 + beta2*r(t-1);

% Conditional variance per state
h1(t-1) = w1 + a1*e^2 + b1*h(t-1);
h2(t-1) = w2 + a2*e^2 + b2*h(t-1);

% Current state and increment of interest rate
u = rand;
if S(t-1) == 1
if u < p11
S(t) = 1;
%             hi(t-1) = h1(t-1);
y(t) = mu1 + sqrt(h1(t-1))*randn;
else
S(t) = 2;
%             hi(t-1) = h2(t-1);
y(t) = mu2 + sqrt(h2(t-1))*randn;
end
elseif S(t-1) == 2
if u < p22
S(t) = 2;
%             hi(t-1) = h2(t-1);
y(t) = mu2 + sqrt(h2(t-1))*randn;
else
S(t) = 1;
%             hi(t-1) = h1(t-1);
y(t) = mu1 + sqrt(h1(t-1))*randn;
end
end

% Interest rate
r(t) = r(t-1) + y(t);

% The loglikelihood function
```

```
g1(t-1)=(1./sqrt(2*pi*h1(t-1))).*exp((-1./(2*h1(t-1))).*(y(t)-mu1).^2);
g2(t-1)=(1./sqrt(2*pi*h2(t-1))).*exp((-1./(2*h2(t-1))).*(y(t)-mu2).^2);


% Regime posterior probabilities
p1(t) = (1-p22)*g2(t-1)*p2(t-1)/(g1(t-1)*p1(t-1) + ...
    g2(t-1)*p2(t-1))...
+ p11*g1(t-1)*p1(t-1)/(g1(t-1)*p1(t-1) + g2(t-1)*p2(t-1));
p2(t) = 1 - p1(t,1);

% Current conditional variances (regime independent)
mu = p1(t)*mu1 + p2(t)*mu2;
e = y(t) - mu;
h(t) = p1(t)*(h1(t-1) + mu1^2) + p2(t)*(h2(t-1) + mu2^2) - ...
    mu^2;

end
SimY_full(:,i)= y(2:end);
SimX_full(:,i)= r(1:end-1);
end
```

C. The function ” OptimiserSimYGRsGarch11Gray.m”

This script on conducts the Monte Carlo study based on the EM algorithm

```
clear; clc; close all;

%% Data
% data = xlsread('SimY_full.xlsx');% data
% data0 =load('SimY_full08282020Gray.mat', 'SimY_full');
% data0X=load('SimY_full08282020Gray.mat', 'SimX_full');

data0 =load('SimY_full08282020FullGray.mat', 'SimY_full');
data0X=load('SimY_full08282020FullGray.mat', 'SimX_full');

data =  data0.SimY_full;
dataX = data0X.SimX_full;

T = length(data);
NbSim =size(data,2);
k1 = 1;
%% Initial values of the parameters

% Regime probabilities
n =2; % number of regimes
P0 =[0.96 0.04;0.1 0.90];
% GARCH parameters
p=1;
q=1;

% param01  = [0.00059; -0.0013; 0.0004 ;0.2095 ; ...
    0.8208;P0(1,1)]; % For regime 1 from Gray Table 3
% param02  = [0.00059; -0.0013; 0.0004 ;0.2095 ; ...
    0.8208;P0(2,2)]; % For regime 2 from Gray Table 3

param01 =[0.25;0.0008;0.5;0.35;0.40;0.9739];
param02 =[-0.07;-0.0006;0.08;0.1;0.2;0.9896];

% param01 =[0.25;0.0008;0.5;0;0;0.9739];
```

```
% param02 =[-0.07;-0.0006;0.08;0;0;0.9896];
parameter00 = [param01 ; param02];
parameter0 = [param01 ; param02];  % Time the number of ...
    regimes, N times
ParamEstimates = zeros(size(parameter0,1)+3,NbSim);
%% Constraintes
nbcon = 12; % number of constraints
A    = zeros(nbcon,size(parameter0,1));
A01 = zeros(nbcon/n,size(param01,1));

A01(1,2+k1) =-1;
A01(2,3+k1) =-1;
A01(3,4+k1) =-1;
A01(4,3+k1:end-1) =1;
A01(5,end) =-1;
A01(6,end) =1;

A(1:nbcon/2,1:size(parameter0,1)/2) = A01;
A(nbcon/2+1:end,size(parameter0,1)/2+1:end) = A01;
% My consttraints are strict inequality but in MatLab the ...
    inequality constrainte is not strict (that<s why I have ...
    substracted -0.0001)
aa = 0.0001;
B01 = [-aa;0;0;1-aa;-aa;1-aa];
% B01 = [-aa;-aa;0;0;1-aa;-aa;1-aa];
B = [B01;B01];

%% Equality constraints
% Aeq   = zeros(nbcon,size(parameter0,1));
% Aeq01 = zeros(nbcon/n,size(param01,1));
% Aeq01(1,2) =1;
% Aeq01(2,end-2) =1;
% Aeq01(3,end-1) =1;
%
% Aeq(1:nbcon/2,1:size(parameter0,1)/2) = Aeq01;
% Aeq(nbcon/2+1:end,size(parameter0,1)/2+1:end) = Aeq01;

nbcon = 4; % number of constraints
Aeq   = zeros(nbcon,size(parameter0,1));
Aeq01 = zeros(nbcon/n,size(param01,1));
```

```

% Aeq   = zeros(4,size(parameter0,1));
% Aeq01 = zeros(2,size(param01,1));
%Aeq01(1,2) =1;
Aeq01(1,end-2) =1;
Aeq01(2,end-1) =1;

Aeq(1:nbcon/2,1:size(parameter0,1)/2) = Aeq01;
Aeq(nbcon/2+1:end,size(parameter0,1)/2+1:end) = Aeq01;

Beq01 = [0;0];
Beq = [Beq01;Beq01];

%% Optimization
%NbSim = 1;

SimPara = zeros(16,NbSim);
TStudent = zeros(12,NbSim);


for i =1:1:NbSim
% for i =1:1:2
Y = data(:,i);
X1 = dataX(:,i);

exitflag0 = 0;
option.MaxFunEvals=10000;
%options = optimset(option,'Display','iter','TolX',1e-6);
options = ...
    optimoptions('fmincon','Display','iter','Algorithm', ...
    'interior-point');
% 'interior-point' 'sqp'
output.firstorderopt0 = 10000;
%fval0=0;
%while output.firstorderopt0 >1e-3

while (exitflag0 ≠1 && exitflag0 ≠2)
f=@(parameter) ...
    RsUnGarch11Gray([parameter(1),parameter(2),parameter(3),parameter(4),paramete
,parameter(7),parameter(8),parameter(9),parameter(10),parameter(11),parameter(12
```

```
[parameter,fval,exitflag,output,lambda,grad,hessian]= ...
    fmincon(f,parameter0,A,B,[],[],[],[],[],options);
std_error = sqrt(diag(pinv(hessian)));
tstat     = (parameter-parameter00)./std_error;
TStudent(1:end,i)= abs(tstat);
ParamEstimates(1:end-3,i)= parameter;
ParamEstimates(end-2,i)= exitflag;
ParamEstimates(end-1,i)= fval;
ParamEstimates(end,i)=output.firstorderopt;

parameter0 = parameter;
output.firstorderopt0 = output.firstorderopt;
exitflag0 = exitflag;
end
end
```

C. The function ”RsUnGarch11Gray.m”

This function evaluates the Maximum Likelihood function of the RSG model of Gray

```
function [LogRsUnGarch11, epsilun,Gi,Hi] = ... GRsUnGarch11Gray(parameter,Y,X1)

% the parameters
k1 = size(X1,2);
N=2;
%np = size(parameter,1)/N;    % Number of parameter to ... estimate by state
np=6;
param =zeros(np,N);

param(1:end,1)=parameter(1:np);
param(1:end,2)=parameter(np+1:end);

c1=param(end-1,1);
d1=param(end,1);

c2=param(end-1,2);
d2=param(end,2);
% % % t=0;
% % % for k=1:N
% % %     param(:,k)=parameter(t+1:k*np,1);
% % %     t=k*np;
% % % end
% parameter  = [alpha ; beta; w; A;B;gamma;P]; % this is a ... (1+k1+3+k2)xN matrix
alpha       = param(1,:);
beta        = param(2:k1+1,:);
% % % w            = exp(param(k1+2,:));
% % % a            = param(k1+3,:).^2;
% % % b            = param(k1+4,:).^2;
w           = param(k1+2,:);
a           = param(k1+3,:);
```

```
b           = param(k1+4,:);

% % % P = param(end,1);
% % % Q = param(end,2);

P = normcdf(c1 +d1*X1);
Q = normcdf(c2 +d2*X1);

%% Initial Values
%---------------------------------------------------------------------------
T = length(Y);
%---------------------------------------------------------------------------
% Initial values for the regime probabilities epsilun
% ...
    -------------------------------------------------------------------------
epsilun = zeros(T+1,N); % It maybe better to take this as a ...
    N*1 vector
epsilun(1,:)=  [(1-Q(1,1))/(2-P(1,1)-Q(1,1)) , ...
    (1-(1-Q(1,1))/(2-P(1,1)-Q(1,1)))];
%---------------------------------------------------------------------------
% The conditional mean mu by regime
%---------------------------------------------------------------------------
mu = zeros(T,N);  % It maybe better to take this as a N*1 ...
    vector
% I compute12d the mu for each obs by regime
for i=1:N
mu(:,i) = ones(T,1)*alpha(i)+X1*beta(:,i);
end

%---------------------------------------------------------------------------
%Computing the conditional variances by regime "hit" for ...
    each period
%---------------------------------------------------------------------------
coefH = [w' a' b'];
Hi = zeros(T+1,N);

Hi(1,:)=var(Y);

% Computing the conditional variance at time t=1
H = zeros(T,1);
```

```
Error =  zeros(T,1);

% The likelihood function by regime
Gi = zeros(T,N);

% computing the H

for t=1:T

H(t) = (mu(t,:).^2+Hi(t,:))*epsilun(t,:)' ...
    -(mu(t,:)*epsilun(t,:)')^2;  %

Error(t) = Y(t)-(mu(t,:)*epsilun(t,:)');  % Y-Xbeta
% The loglikelihood function
Gi(t,:)=(1./sqrt(2*pi*Hi(t,:))).*exp((-0.5*(Y(t)-mu(t,:)).^2)./Hi(t,:));

% the next period regime probabilities
epsilun(t+1,1)= ...
    (1-Q(t,1))+((P(t,1)+Q(t,1)-1)*(Gi(t,1)*epsilun(t,1))/(Gi(t,:)*epsilun(t,:)'))
epsilun(t+1,2)= 1-epsilun(t+1,1);
% Hi
Hi(t+1,:) =[1  Error(t)*Error(t)  H(t)]*coefH' ;
end
% Deleting the additional row of the regimes probabilities ...
    matrix
epsilun=epsilun(1:T,:);% X2(T+1,:) =[]; %Hadd(T+1,:) =[];

% the log likelihood is to be maximzed but bcz I am using ...
    fmincon I will
% multiply by -1
LogRsUnGarch11 = -sum(log((epsilun.*Gi)*ones(N,1)));

LogRsUnGarch11(isnan(LogRsUnGarch11))=1e6;

LogRsUnGarch11(isinf(LogRsUnGarch11))=1e6;
```