Università
Ca' Foscari

Venezia

Dottorato di ricerca
in Economia
Scuola di dottorato in Economia
Ciclo XXV
(A.A. 2013-2014)

Essays on Bayesian inference with financial applications

settore scientifico disciplinare di afferenza: SECS-P05

Tesi di dottorato di **OSUNTUYI AYOKUNLE ANTHONY** , Matricola 955710

Coordinatore del Dottorato
Prof. Michele Bernasconi

Tutore del dottorando
Prof. Monica Billio

Co-tutore del dottorando
Prof. Roberto Casarin

The undersigned OSUNTUYI AYOKUNLE ANTHONY, in his quality of doctoral candidate for a Ph.D. degree in Economics granted by the Università Ca' Foscari Venezia attests that the research exposed in this dissertation is original and that it has not been and it will not be used to pursue or attain any other academic degree of any level at any other academic institution, be it foreign or Italian.

Abstract

This thesis is composed of two main research lines. The first line, developed in chapters 2 to 4, deals with frequentist and Bayesian estimation of regime-switching GARCH models and its application to risk management on energy markets, while the second part, which corresponds to chapter 5, focuses on forecast rationality testing within a Bayesian framework.

Chapter 2 presents a unified mathematical framework for characterizing the class of MSGARCH models based on collapsing the regimes in order to eliminate the usual path dependence problem. Within this framework, two new models (identified as Basic model and Simplified Klaassen model) are proposed as alternative specifications of the MS-GARCH model. Using Maximum Likelihood Estimation, we estimate the parameters of the different models within this family and compare their performance on both simulation and empirical exercises. Chapter 3 proposes new efficient Monte Carlo simulation techniques based on multiple proposal Metropolis. The application to approximated inference for regime-switching GARCH models is there discussed. In Chapter 4, we provide an extension of our efficient Monte Carlo simulation algorithm to a multi-chain Markov switching multivariate GARCH model and apply it to risk management in commodity market. More specifically we focus on futures commodity market and suggest a dynamic and robust minimum variance hedging strategy which accounts for model parameter uncertainty. In chapter 5, we propose a new Bayesian inference procedure for testing the monotonicity properties of second moment bounds across several horizons presented in Patton and Timmermann [2012].

Acknowledgements

My profound gratitude goes to my PhD advisors, Monica Billio and Roberto Casarin, for mentoring me in research, critical thinking, and scientific writing. I am indebted to their invaluable insight, suggestion for improvements, and motivation during the program. I also wish to thank Lennart Hoogerheide, and Herman K. van Dijk for their contribution to my thesis and for their hospitality during my research visit to Tinbergen Institute, Vrije Universiteit Amsterdam, The Netherlands.

I am grateful to my parents for their love and unceasing support during my graduate studies. Also, I am thankful to my wife, Olawumi Osuntuyi, for her encouragement and understanding when the going was touugh.

I would also like to express my gratitude to the management of Obafemi Awolowo University (O.A.U.), Ile-Ife, Nigeria for granting me the opportunity to take part in this doctoral programme. The support and understanding of members of staff of the Department of Mathematics, O.A.U, Ile-Ife is also duly acknowledged.

My thanks won't be complete without expressing profound gratitude: To God for granting me good health and for making this feat a reality; To my colleagues for making my stay a memorable one; To Drs. Shola Adeyemi, Sajojbi Tolulope, Fareo Gideon, Ireka Ikenna, Owatemi Dare for inspiring me to always strive for the best.

To my loving parents, wife and son.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The first part of this research work addresses the estimation of Markov-switching generalized autoregressive conditional heteroscedasticity (MS-GARCH) models and their applications to risk management, while the second part focuses on forecast rationality tests.

Volatility modeling has remained an active area of research in finance because of the important role it plays in a variety of financial problems. For example, portfolio theory, asset pricing, and hedging all involve volatilities. Since the introduction of the generalized autoregressive conditional heteroscedasticity (GARCH) model by Bollerslev [1986], a remarkable number of research papers providing extensions and application of this modeling strategy to financial time series data have emerged in the literature. These extensions recognize that there may be important nonlinearity, asymmetry, and long memory properties in the volatility process. However, empirical studies have shown that shocks to the volatility process described by a single regime GARCH model exhibit high persistence when measured using estimates from daily data. Based on this empirical observation, early studies have accounted for this problem by considering integrated GARCH process (see Engle and Bollerslev [1986] and Baillie and Mikkelsen [1996] for illustration). In recent years, argument in favour of model mis-specification has been advanced in the literature to be responsible for the high persistence problem. See Lamoureux and Lastrapes [1990] and Mikosch and Starica [2004]. Following this finding, there has been a shift of interest to regime switching GARCH models.

A Markov Switching GARCH (MS-GARCH) model is a nonlinear specification of the evolution of a time series assessed to be affected by different states of the world and for which the conditional variance in each state follow a GARCH process. Inherent in the MS-GARCH framework is the well known path dependence problem which renders its estimation via maximum likelihood procedure a daunting task. To tackle this estimation problem via maximum likelihood procedure, procedures based on model approximation have been suggested in the literature. See Gray [1996], Dueker [1997], and Klaassen [2002] for illustration. Apart from the models suggested in the literature, new auxiliary MS-GARCH models may still be proposed and this represents a possible area of research.

Although the various approximations to the estimation problem of the regime switching GARCH models has proven to be a success, their analytical intractability is a major drawback i.e. they cannot be verified by any of the well known analytical approximation methods available. On this premise and thanks to the advent of computers with high computing capabilities, new and extant statistical and mathematical methods such as simulated maximum likelihood (Augustyniak [2013]), Bayesian methods (Bauwens et al. [2010]), particle filter (Bauwens et al. [2011]) and neural networks (Bildirici and Ersin [2012]), have been been proposed and applied as alternative estimation technique for MS-GARCH models. However, most of these methods as applied in estimating the MS-GARCH model are known to be either computationally intensive or costly to code because of their complexity. This situation therefore suggests that for practical purposes more research work is required in this area in order to provide efficient estimators that can compete favourably with existing methods.

In the hedging literature, MS-GARCH models are often used for characterizing the dynamics of both the spot and derivative returns. See Haigh and Holt [2002] and Lee and Yoder [2007]. Most applications of MS-GARCH models are based on the assumption that both the spot and derivative are driven by the same hidden state process. Arguably, this assumption may not hold true since there exists the possibility that spot and futures returns might follow different switching dynamics governed by different state variables. See Sheu and Lee [2012] and references there in for further discussion on this issue. In line with this argument, Sheu and Lee [2012] propose the use of multichain Markov regime switching GARCH (MCSG) models. More specifically, the authors assume that the state dependent variances of spot and futures returns are driven by independent state variables while the correlation between the two variables depends on both state variables. To circumvent the path dependence problem and facilitate the estimation via maximum likelihood procedure, the authors apply the recombining method of Gray [1996]. From a modeling point of view, the use of the multivariate GARCH framework as applied in most hedging literature requires a definite specification of the dynamics of the covariance matrix. Some of these specifications include the Matrix-Diagonal (MD) GARCH, the BEKK-GARCH and the Constant Conditional Correlation (CCC) GARCH. The use of these specifications may have different implication on the hedge ratio. An alternative modeling framework which avoids a direct specification of the dynamics of the covariance matrix is to consider a simultaneous modeling of the return dynamics on the hedged portfolio and the futures. The reduced form representation of the simultaneous system corresponds to a MS-GARCH model with covariance matrix different from the ones considered in the literature. The impact of combining the multichain framework with the simultaneous equation on the optimal hedge ratio will surely provide an interesting study.

Digressing a little from volatility modeling and estimation, the second part of this research work is concerned with forecasting. Forecasts are used by policymakers and other practitioners for making informed decisions in their businesses. For example, Government's decision on taxes, revenues, money and credit supply, foreign trade and balances, employment among others rests upon actual or implied forecasts of economic conditions. In finance, on the other hand, forecast plays an important role in such studies as portfolio theory, asset pricing and corporate finance among others. Research work in forecasting may be grouped into two major areas: (i) forecast construction and (ii) forecasts evaluation. Although case (i) is a relevant area of research interest (Elliot and Timmermann [2004], Geweke and Whiteman [2006], Hashem Pesaran et al. [2013]), we are here concerned only with case (ii).

Forecasts on a single economic or financial variable are usually published by several independent forecasters and, generally, the underlying forecasting processes used for generating these forecasts are not made known. Hence, judging whether these available forecasts are good or not is vital in the decision making process of end users. Typical forecast rationality testing procedures are constructed for forecast with one horizon (Zellner [1986], Patton and Timmermann [2007], Aretz et al. [2011]). However, forecast on economic and financial variables are usually reported on a multi-horizon basis. The availability of such forecasts provides an opportunity to develop tests of optimality that exploit the information in the complete "term structure" of forecasts recorded across all horizons. Rather than focusing separately on individual horizon, to exploit the information across several horizons by the use of multi-horizon forecast tests offers the potential of drawing more powerful conclusions about the ability of forecasters to produce optimal forecasts. Patton and Timmermann [2012] and Capistran [2007] propose several test procedures following this argument. While the Patton and Timmermann [2012] rationality test has brought a new dimension to the evaluation of forecasts, the procedure relies heavily on asymptotic distributions under the null hypothesis and consequently to be able to produce more reliable results for moderate and small sample data needs adequate investigation. The Bayesian framework represents an alternative tool that for addressing this issue.

1.1 Purpose and Objectives

The aim of this research is to contribute to the statistical analysis of financial market data. The objectives are to develop:

1. new auxiliary MS-GARCH models;
2. new efficient methods for estimating MS-GARCH models;
3. new model for hedging and new provide efficient estimation procedure;
4. new Bayesian inference procedure for testing the monotonicity properties of second moment bounds across several horizons presented in Patton and Timmermann [2012].

1.2 Organization of Thesis

This thesis is organized into two parts composed of four manuscripts. Part 1 begins with Manuscript 1, *A Unified class of Markov switching GARCH models based on Collapsing procedure*, which is presented in Chapter 2. This manuscript reviews the literature on MS-GARCH models and develops a family of MS-GARCH processes that not only encompasses some of the specifications introduced in the literature but also identifies new ones. This class of models shares a common feature of eliminating the path dependence problem by "collapsing" the history of the regimes in some way i.e. the conditional variance of being in the current regime is dependent on the expectation of the previous conditional variances rather than their values. Based on the concern related to the unverifiable nature of these approximations, Manuscript 2, *Efficient Gibbs Sampling for Markov Switching GARCH Models*, presented in Chapter 3, provides an alternative estimation strategy for the MS-GARCH model. In this manuscript, new efficient Monte Carlo simulation techniques based on multiple proposal Metropolis are proposed. The application to approximated inference for regime-switching GARCH models is there discussed. Manuscript 3, *Markov Switching GARCH models for Bayesian Hedging on Energy Futures Markets*, presented in Chapter 4, extends the application of the methodology developed in Chapter 3 to a multichain multivariate GARCH framework. The application to futures hedging is then undertaken.

The second part of this thesis corresponding to Chapter 5 focuses on forecast rationality tests. This chapter which contains Manuscript 4, *Bayesian Approach to Forecast Rationality Tests of A.J. Patton and A. Timmermann*, a joint work with Lennart Hoogerheide, and Herman K. van Dijk, proposes a new Bayesian inference to forecast rationality tests. These rationality tests are based on the monotonicity properties of second moment bounds across several horizons as identified in Patton and Timmermann [2012]. This approach avoids the use of computationally expensive test statistics and asymptotic distributions under the null hypothesis and consequently produces more reliable results for moderate and small sample data. We are also able to account for parameter uncertainty using our proposed Bayesian inference technique.

Bibliography

Kevin Aretz, Söhnke M Bartram, and Peter F Pope. Asymmetric loss functions and the rationality of expected stock returns. *International Journal of Forecasting*, 27(2):413–437, 2011. URL http://www.sciencedirect.com/science/article/pii/S0169207009001769.

M. Augustyniak. Maximum likelihood estimation of the Markov-switching GARCH model. *Computational Statistics & Data Analysis*, 2013.

Bollerslev T. Baillie, R. T. and H. O. Mikkelsen. Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of econometrics*, 74(1):3–30., 1996.

L. Bauwens, A. Preminger, and J. Rombouts. Theory and inference for a Markov switching GARCH model. *Econometrics Journal*, 13:218–244, 2010.

L. Bauwens, A. Dufays, and J. Rombouts. Marginal likelihood for Markov-switching and change-point GARCH. *CORE discussion paper, 2011/13*, 2011.

M. Bildirici and O. Ersin. Modeling markov switching arma-garch neural networks models and an application to forecasting stock returns. *Available at SSRN 2125855.*, 2012.

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

C. Capistran. Optimality Tests for Multi-Horizon Forecast. Working Paper 2007-14, Banco de Mexico, 2007.

M. Dueker. Markov switching in GARCH processes in mean reverting stock market volatility. *Journal of Business and Economics Statistics*, 15:26–34, 1997.

G. Elliot and A. Timmermann. Optimal Forecast Combination Under General Loss Function and Forecast Error Distributions. *Journal of Econometric*, 122:47 – 79, 2004.

R. F. Engle and T. Bollerslev. Modelling the persistence of conditional variances. *Econometric review*, 5(1):1–50, 1986.

J. Geweke and C. Whiteman. Bayesian Forecasting. In G. Elliott, C.W.J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, pages 3 – 80. Elsevier, 2006.

S. F. Gray. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42:27–62, 1996.

M. S. Haigh and M. T. Holt. Crack spread hedging: accounting for time-varying volatility spillovers in the energy futures markets. *Journal of Applied Econometrics*, 17:269 – 289, 2002.

M Hashem Pesaran, Andreas Pick, and Mikhail Pranovich. Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 2013. URL http://www.sciencedirect.com/science/article/pii/S0304407613000687.

F. Klaassen. Improving GARCH volatility forecasts with regime switching GARCH. *Empirical Economics*, 27:363–394, 2002.

C. Lamoureux and W. Lastrapes. Persistence in variance, structural change, and the GARCH model. *Journal of Business and Economics Statistics*, 8:225–234, 1990.

H. Lee and J. K. Yoder. A bivariate Markov regime switching GARCH approach to estimate time varying minimum variance hedge ratio. *Applied Economics*, 39(10): 1253 – 1265, 2007.

T. Mikosch and C. Starica. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86: 378–390, 2004.

A. J. Patton and A. Timmermann. Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity. *Journal of Econometric*, 140:884 – 918, 2007.

A.J. Patton and A. Timmermann. Forecast rationality tests based on multi-horizon bounds. *Journal of Business & Economic Statistics*, 30(1):1–17, 2012.

H.-J. Sheu and H.-T. Lee. Optimal futures hedging under multichain markov regime switching. *Journal of Futures Markets*, 2012.

Arnold Zellner. Biased predictors, rationality and the evaluation of forecasts. *Economics Letters*, 21(1):45–48, 1986. URL http://www.sciencedirect.com/science/article/pii/0165176586901199.

Chapter 2

A Unified class of Markov switching GARCH models based on Collapsing procedure

Abstract : We present a unified mathematical framework for MS-GARCH models which enables us (1) to easily identify the functional form of various MS-GARCH models based on regime collapsing procedure as available in the literature, and (2) to propose new variants of the MS-GARCH models. Using this framework, two new models tagged Basic model and Simplified Klaassen model are proposed as alternative specifications of the MS-GARCH model. Maximum Likelihood Estimation (MLE) procedure is then used to estimate the parameters of the different models belonging to this family and a comparison of their performances is conducted on both simulated and empirical data.

Keywords : GARCH, Markov switching, Volatility, MLE

2.1 Introduction

In the last two decades, a lot of research has been devoted, both theoretically and empirically, to modelling volatility of financial markets. This is because volatility plays an important role in understanding and addressing financial problems such as asset pricing and and financial risk management.

Bollerslev [1986] Generalized Autoregressive Conditional Heteroskedastic (GARCH) model ranks as the most popular class of volatility models among researchers and practitioners (e.g., see Ardia [2008]) as it is able to capture some of the well known stylized facts (e.g., heavy tails, volatility clustering, no correlation in returns but correlation in the squares e.t.c.) of financial data. A simple GARCH(1,1) model has the following representation:

$$
\begin{aligned}
y_t &= \mu + \sigma_t \eta_t \qquad \eta_t \overset{iid}{\sim} (0,1) \\
\sigma_t^2 &= \gamma + \alpha \epsilon_{t-1}^2 + \beta \sigma_{t-1}^2 \\
\epsilon_t &= y_t - \mu = \sigma_t \eta_t
\end{aligned}
$$

where y_t represents returns on some financial asset, μ represents the parameters of the conditional mean of the returns series, $\gamma > 0$, $\alpha \geq 0$, $\beta \geq 0$, are the GARCH parameters. To ensure covariance stationarity of the process it is required that $\alpha + \beta << 1$.

Several extensions and refinements of the GARCH model have been proposed to account for additional stylized facts observed in financial markets. These extensions recognize that there may be important nonlinearity, asymmetry, and long memory properties in the volatility process. A review of these models can be found Bollerslev et al. [1992], Bollerslev and Engle [1994], Engle [2004]. Prominent ones among these models are the Exponential GARCH model by Nelson [1991] and the GJR model by

Glosten et al. [1993], both account for the asymmetric relation between stock returns and changes in variance (see Black [1976]).

As appealing as this class of models is, the shocks to the volatility process are known to exhibit high persistence when measured using estimates from daily data. For example in Equation 2.1, the measure of volatility persistence $\alpha + \beta$, when estimated from daily data is approximately 1. This is not a desirable property as it violates the covariance stationarity assumption. Based on this observation, Engle and Bollerslev [1986] propose integrated GARCH process which take $\alpha + \beta = 1$, and prevents the shocks on the volatility from dying out over time. However, Lamoureux and Lastrapes [1990], among others, argue that the presence of structural changes in the variance process, for which the standard GARCH process cannot account for, may be responsible for this phenomenon. To buttress this point, Mikosch and Starica [2004] estimate a GARCH model on a sample that exhibits structural changes in its conditional variance and obtained a nearly integrated GARCH effect from the estimate. Based on this finding, it is instructional for practical purposes to assume a regime-switching volatility models when handling financial problem involving volatility.

The class of MS-GARCH models is gradually becoming a work house among economics and financial practitioners for analysing financial markets data (e.g., see Marcucci [2005]). For practical implementation of this class of theoretical models, it is crucial to have reliable parameter estimators. Maximum Likelihood (ML) approach is a natural route to parameter estimation in Econometrics. However, the ML technique is not computationally feasible for MS-GARCH models because of the path dependence problem (see Gray [1996]). To facilitate estimation via ML, Cai [1994] and Hamilton and Susmel [1994] considered a MS-ARCH model. This approach effectively makes the model tractable because the lagged conditional variance that makes the conditional variance dependent on the history of regime has been dropped. Another approach propose in the literature is to modify MS-GARCH model such that the path dependent problem can be eliminated. For example, Gray [1996] noted that the conditional density of the return is essentially a mixture of distributions with time-varying mixing parameter and in particular under normality assumption he suggested the use of aggregate conditional variances over all regimes as the lagged conditional variance when constructing the conditional variance at each time step. Extensions of Gray [1996] model can be found in Dueker [1997] and Klaassen [2002] among others. Abramson and Cohen [2007] provide stationarity conditions for some of these approximations. The collapsing procedure has been shown to be easy to implement and effective. Other estimation methods such as simulated maximum likelihood procedure (Elliott et al. [2012], Augustyniak [2013]), Generalized method of moments (GMM) procedure (Francq et al. [2001]), neural networks (Bildirici and Ersin [2012]) and Bayesian Markov chain Monte Carlo (MCMC) algorithm (He and Maheu [2010], Bauwens et al. [2010], Bauwens et al. [2011], Dufays [2012]) have been proposed in the literature.

The purpose of this chapter is to develop a framework for a family of MS-GARCH processes encompassing some of the existing models in the literature, such as Gray [1996], Dueker [1997] and Klaassen [2002], and also allowing for the identification new variants of the MS-GARCH model. This class of models is developed by replacing the lagged conditional variance in the variance process with an average. Furthermore, the information set upon which the expectation is taken differentiates the models belonging to this class. This class of models are known to be less computationally expensive and effective. Liu [2012] on the other hand develop a different class of the MS-GARCH processes by exploiting the common feature shared by Markov-switching and GARCH processes and applying a similar approach taken by He and Teräsvirta [1999]. The nesting of the models in this class relies on a Box and Cox [1964] transformation with shape parameter (d > 0) to the conditional variance process and on a response function driven by Markov chain to shocks.

This chapter is organized as follows. Section 2.2 defines the MS-GARCH model. Section 2.3 introduces the new class of MS-GARCH models. In section 2.4, a simulation exercise is carried out to demonstrate the effectiveness of the models in this class. In section 2.5, the Klaassen's model is fitted to daily log-returns on the S&P 500 index. Section 2.6 concludes the paper.

2.2 Markov switching GARCH models

A Markov Switching GARCH model is a nonlinear specification of the evolution of a time series assessed to be affected by different states of the world and for which the conditional variance in each state follows a GARCH process. More specifically, let y_t be the observed variable (e.g. the return on some financial asset) and s_t a discrete, unobserved, state variable which could be interpreted as the state of the world at time t. Define $(y_t, \ldots, y_u)$ and $(s_t, \ldots, s_u)$ as $y_{t:u}$ and $s_{t:u}$ respectively whenever $t \leq u$. Then

$$y_t = \mu_t(y_{1:t-1}, \theta_\mu(s_t)) + \sigma_t(y_{1:t-1}, \theta_\sigma(s_t))\eta_t, \qquad \eta_t \overset{iid}{\sim} N(0,1), \tag{2.1}$$

$$\sigma_t^2(y_{1:t-1}, \theta_\sigma(s_t)) = \gamma(s_t) + \alpha(s_t)\epsilon_{t-1}^2 + \beta(s_t)\sigma_{t-1}^2(y_{1:t-2}, \theta_\sigma(s_{t-1})), \tag{2.2}$$

where, $N(0,1)$ denotes the standard normal distribution, θ_μ represents the parameters of the conditional mean of the returns series, $\epsilon_t = \sigma_t(y_{1:t-1}, \theta_\sigma(s_t))\eta_t$, $\theta_\sigma(s_t) = (\gamma(s_t), \alpha(s_t), \beta(s_t))$, $\gamma(s_t) > 0$, $\alpha(s_t) \geq 0$, $\beta(s_t) \geq 0$, and $s_t \in \{1, \ldots, M\}$, $t = 1, \ldots, T$, is assumed to follow a M-state first order Markov chain with transition probabilities $\{\pi_{ij}\}_{i,j=1,2,\ldots,M}$:

$$\pi_{ij} = p(s_t = i | s_{t-1} = j, \theta_\pi), \quad \sum_{i=1}^{M} \pi_{ij} = 1 \quad \forall\ j = 1, 2, \ldots, M,$$

where θ_π represents the parameters of the transition matrix. The parameter shift functions $\gamma(s_t)$, $\alpha(s_t)$ and $\beta(s_t)$, describe the dependence of parameters on the regime s_t i.e.

$$\gamma(s_t) = \sum_{m=1}^{M} \gamma_m \mathbb{I}_{s_t=m}, \quad \alpha(s_t) = \sum_{m=1}^{M} \alpha_m \mathbb{I}_{s_t=m}, \quad \text{and} \quad \beta(s_t) = \sum_{m=1}^{M} \beta_m \mathbb{I}_{s_t=m},$$

where,

$$\mathbb{I}_{s_t=m} = \begin{cases} 1, & \text{if } s_t = m, \\ 0, & \text{otherwise.} \end{cases}$$

By defining the allocation variable, s_t, as a M-dimensional discrete vector, $\xi_t = (\xi_{1t}, \ldots, \xi_{Mt})'$, where $\xi_{mt} = \mathbb{I}_{s_t=m}$, $m = 1, \ldots, M$, the system of equations in (2.1)-(2.2) can be written compactly as

$$y_t = \mu_t(y_{1:t-1}, \xi_t'\theta_\mu) + \sigma_t(y_{1:t-1}, \xi_t'\theta_\sigma)\eta_t, \qquad \eta_t \overset{iid}{\sim} N(0,1), \tag{2.3}$$

$$\sigma_t^2(y_{1:t-1}, \xi_t'\theta_\sigma) = (\xi_t'\gamma) + (\xi_t'\alpha)\epsilon_{t-1}^2 + (\xi_t'\beta)\sigma_{t-1}^2(y_{1:t-2}, \xi_{t-1}'\theta_\sigma), \tag{2.4}$$

where $\epsilon_t = \sigma_t(y_{1:t-1}, \xi_t'\theta_\sigma)\eta_t$, $\gamma = (\gamma_1, \ldots, \gamma_M)'$, $\alpha = (\alpha_1, \ldots, \alpha_M)'$, $\beta = (\beta_1, \ldots, \beta_M)'$, $\theta_\mu = (\theta_{1\mu}, \ldots, \theta_{M\mu})'$ and $\theta_\sigma = (\theta_{1\sigma}, \ldots, \theta_{M\sigma})'$ with $\theta_{m\sigma} = (\gamma_m, \alpha_m, \beta_m)'$ for $m = 1, \ldots, M$. for $t = 1, \ldots, T$. Let $\pi = (\pi_1, \ldots, \pi_M)$, with $\pi_i = (\pi_{i1}, \ldots, \pi_{iM})$ for $i = 1, 2, \ldots, M$ and $\sum_{i=1}^{M} \pi_{ij} = 1$ for all $j = 1, 2, \ldots, M$. Since ξ_t follows a $M-$state first order Markov chain, we define the transition probabilities $\{\pi_{ij}\}_{i,j=1,2,\ldots,M}$ by

$$\pi_{ij} = p(\xi_t = e_i | \xi_{t-1} = e_j, \theta_\pi),$$

where e_i is the $i-$th column of a M-by-M identity matrix. The conditional probability of ξ_t given ξ_{t-1}, θ_π and $y_{1:t-1}$ is given by

$$p(\xi_t|\xi_{t-1}, \theta_\pi) = \prod_{m=1}^{M} (\pi_m \xi_{t-1})^{\xi_{mt}}, \tag{2.5}$$

which implies that the probability with which event m occurs at time t is $\pi_m \xi_{t-1}$.

2.2.1 Path dependence issue

Estimating Markov switching GARCH models is a challenging problem since the likelihood of y_t depends on the entire sequence of past states up to time t due to the recursive structure of its volatility. To elaborate on this, the likelihood function of the switching GARCH model is given by

$$\mathcal{L}(\theta|y_{1:T}) \equiv f(y_{1:T}|\theta) = \sum_{i_1=1}^{M} \cdots \sum_{i_T=1}^{M} f(y_{1:T}, \xi_1 = e_{i_1}, \ldots, \xi_T = e_{i_T}|\theta) \tag{2.6}$$

where $\theta = (\{\theta_{m\mu}, \theta_{m\sigma}\}_{m=1,\ldots,M}, \theta_\pi)$. Setting $\xi_{s:t} = (\xi'_s, \ldots, \xi'_t)$ whenever $s \leq t$, the joint density function of $y_{1:t}$ and $\xi_{1:t}$ on the right hand side of equation (2.6) is

$$\begin{aligned} f(y_{1:T}, \xi_{1:T}|\theta) &= f(y_1|\xi_{1:1}, \theta_\mu, \theta_\sigma) \prod_{t=2}^{T} f(y_t|y_{1:t-1}, \xi_{1:t}, \theta_\mu, \theta_\sigma) p(\xi_t|\xi_{1:t-1}, \theta_\pi) \\ &= f(y_1|\xi_{1:1}, \theta_\mu, \theta_\sigma) \prod_{t=2}^{T} f(y_t|y_{1:t-1}, \xi_{1:t}, \theta_\mu, \theta_\sigma) \left(\prod_{i=1}^{M} (\pi_i \xi_{t-1})^{\xi_{it}} \right), \end{aligned} \tag{2.7}$$

with,

$$f(y_t|y_{1:t-1}, \xi_{1:t}, \theta_\mu, \theta_\sigma) \propto \frac{1}{\sigma_t(y_{1:t-1}, \xi'_t \theta_\sigma)} \exp\left(-\frac{1}{2} \left(\frac{y_t - \mu_t(y_{1:t-1}, \xi'_t \theta_\mu)}{\sigma_t(y_{1:t-1}, \xi'_t \theta_\sigma)} \right)^2 \right).$$

Given σ_1, recursive substitution in equation (2.4) yields

$$\sigma_t^2 = \sum_{i=0}^{t-2} \left[\xi'_{t-i}\gamma + (\xi'_{t-i}\alpha)\epsilon_{t-1-i}^2\right] \prod_{j=0}^{i-1} \xi'_{t-j}\beta + \sigma_1^2 \prod_{i=0}^{t-2} \xi'_{t-i}\beta. \tag{2.8}$$

Equation (2.8) clearly shows the dependence of conditional variance at time t on the entire history of the regimes and thus the dependence of the likelihood function on the entire history of the regimes. The evaluation of the likelihood function over a sample of length T, as can be seen in equation (2.6), involves integration (summation) over all M^T unobserved states i.e. integration over all M^T possible (unobserved) regime paths. This requirement makes the maximum likelihood estimation of θ infeasible in practice.

In order to tackle this estimation problem via maximum likelihood procedure, procedures based on model approximation have been suggested in the literature. This approach has been shown (see Gray [1996], Dueker [1997], Klaassen [2002] for illustration) to be very convenient, easy to implement and effective. In the next section we develop a family of MS-GARCH models based on model approximation.

2.3 A family of MS-GARCH models

A possible way of circumventing the path dependence problem inherent in the MS-GARCH model is to replace the lagged conditional variance appearing in the definition of the GARCH model with a proxy. In general, we consider the following class of MS-GARCH(1,1) model

$$y_t = \mu_t(y_{1:t-1}, \xi'_t\theta_\mu) + \sigma_t(y_{1:t-1}, \xi'_t\theta_\sigma)\eta_t, \qquad \eta_t \overset{iid}{\sim} N(0,1), \tag{2.9}$$

$$\sigma_t^2 = (\xi'_t\gamma) + (\xi'_{t_1}\alpha)\epsilon_{(X)t-1}^2 + (\xi'_{t_2}\beta)\sigma_{(X)t-1}^2, \tag{2.10}$$

where $t_1, t_2 \leq t$,

$$\epsilon_{(X)t-1} = y_{t-1} - E[\mu_{t-1}(y_{1:t-2}, \xi_{t-1}^{'}\theta_\mu)|\mathcal{I}_{t-1}]$$
$$\sigma^2_{(X)t-1} = E[\sigma^2_{t-1}|\mathcal{I}_{t-1}] + V(\mu_{t-1}(y_{1:t-2}, \xi_{t-1}^{'}\theta_\mu)|\mathcal{I}_{t-1}),$$

and $\mathcal{I}_{t-1}$ is a information set at time $t-1$. The parameters and transition probabilities remains as defined in (2.2). Alternative specification of $\epsilon_{(X)t-1}$ and $\sigma^2_{(X)t-1}$ amounts to different MS-GARCH models. The variable X is a label that takes up various values such as B, G, D, K, KL, representing different approximations. This new family includes:

1. Gray's approximation (model G): Gray [1996] model is obtained by setting $t_1 = t_2 = t$ and $\mathcal{I}_{t-1} = \{y_{1:t-2}\}$ i.e.

$$\sigma_t^2 = (\xi_t'\gamma) + (\xi_t'\alpha)\epsilon^2_{(G)t-1} + (\xi_t'\beta)\sigma^2_{(G)t-1},$$

with

$$\epsilon_{(G)t-1} = y_{t-1} - \mu_{(G)t-1}$$
$$\mu_{(G)t-1} = E[\mu_{t-1}(y_{1:t-2}, \xi_{t-1}'\theta_\mu)|y_{1:t-2}]$$
$$\sigma^2_{(G)t-1} = V(\mu_{t-1}(y_{1:t-2}, \xi_{t-1}'\theta_\mu)|y_{1:t-2}) + E[\sigma^2_{t-1}(y_{1:t-2}, \xi_{t-1}'\theta_\sigma)|y_{1:t-2}]$$

2. Dueker's approximation (model D): Dueker [1997] model may be obtained by setting $t_1 = t_2 = t-1$ and $\mathcal{I}_{t-1} = \{y_{1:t-1}\}$ i.e.

$$\sigma_t^2 = (\xi_t'\gamma) + (\xi_{t-1}'\alpha)\epsilon^2_{(D)t-1} + (\xi_{t-1}'\beta)\sigma^2_{(D)t-1}(\xi_{t-1}^{'}),$$

with

$$\epsilon_{(D)t-1} = y_{t-1} - \mu_{(D)t-1}$$

$$\mu_{(D)t-1} = E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-1}]$$

$$\sigma^2_{(D)t-1} = E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-2}\theta_\sigma)|y_{1:t-1}].$$

3. Klaassen's approximation (model K): Klaassen [2002] MS-GARCH(1,1) model is obtained by setting $t_1 = t_2 = t$ and $\mathcal{I}_{t-1} = \{y_{1:t-1}, \xi_t\}$ i.e.

$$\sigma_t^2 = (\xi_t'\gamma) + (\xi_t'\alpha)\epsilon^2_{(K)t-1} + (\xi_t'\beta)\sigma^2_{(K)t-1}(\xi_t^{'}),$$

with

$$\epsilon_{(K)t-1} = y_{t-1} - \mu_{i,(K)t-1}$$

$$\mu_{i,(K)t-1} = E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-1}, \xi_t = e_i]$$

$$\sigma^2_{i,(K)t-1} = E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-1}, \xi'_{t-2})|y_{1:t-1}, \xi_t = e_i].$$

In addition, the system of Equation (2.9)-(2.10) also give many new MS-GARCH models. For example, we identify the following new MS-GARCH models and tag them Basic and Simplified Klaassen approximation.

1. Basic approximation (model B): this is obtained by setting $t_1 = t_2 = t$ and $\mathcal{I}_{t-1} = \{y_{1:t-2}\}$ i.e.

$$\sigma_t^2 = (\xi_t'\gamma) + (\xi_t'\alpha)\epsilon^2_{(B)t-1} + (\xi_t'\beta)\sigma^2_{(B)t-1},$$

with

$$
\begin{aligned}
\epsilon_{(B)t-1} &= y_{t-1} - \mu_{(B)t-1} \\
\mu_{(B)t-1} &= E[\mu_{t-1}(y_{1:t-2}, \xi_{t-1}'\theta_\mu)|y_{1:t-2}] = E[y_{t-1}|y_{1:t-2}] \\
&= \sum_{m=1}^{M} \mu_{t-1}(y_{1:t-2}, e_m'\theta_\mu) q(\xi_{t-1} = e_m|y_{1:t-2}), \\
\sigma^2_{(B)t-1} &= E[\sigma^2_{t-1}(y_{1:t-2}, \xi_{t-1}'\theta_\sigma)|y_{1:t-2}] = E[\epsilon^2_{t-1}|y_{1:t-2}] = V(\epsilon_{t-1}|y_{1:t-2}) \\
&= \sum_{m=1}^{M} \sigma^2_{t-1}(y_{1:t-2}, e_m'\theta_\sigma) q(\xi_{t-1} = e_m|y_{1:t-2}).
\end{aligned}
$$

If $\mu_t = 0 \ \forall \ t$, then this approximation is equivalent to Gray's approximation. Observe that in this approximation scheme $\mu_{(B)t-1}$ and $\sigma^2_{(B)t-1}$ are functions of $y_{1:t-2}$ and the information coming from y_{t-1} is lost. In line with Gray's approach, $\sigma^2_{(B)t-1}$ is equal to the variance of the conditional density of ϵ_t. For practical implementation of this approximation, given $q(\xi_{t-1} = e_m|y_{1:t-2})$ for $m = 1, \ldots, M$, $\mu_{(B)t-1}$ can easily be computed while $\sigma^2_{(B)t-1}$ can be computed recursively since $\sigma^2_{t-1}(y_{1:t-2}, e_m'\theta_\sigma)$ depends on $\sigma^2_{(B)t-2}$. This auxiliary model uses the least information, $y_{1:t-2}$, among the auxiliary models under consideration. Hence, we tag it basic model approximation.

2. Simplified Klaassen approximation (model SK) is obtained by letting $t_1 = t_2 = t$ and $\mathcal{I}_{t-1} = \{y_{1:t-1}\}$ i.e.

$$
\sigma^2_t = (\xi_t'\gamma) + (\xi_t'\alpha)\epsilon^2_{(SK)t-1} + (\xi_t'\beta)\sigma^2_{(SK)t-1},
$$

with

$$\epsilon_{(SK)t-1} = y_{t-1} - \mu_{(SK)t-1}$$

$$\mu_{(SK)t-1} = E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-1}] = \sum_{m=1}^{M} \mu_{t-1}(y_{1:t-2}, e'_m\theta_\mu)q(\xi_{t-1} = e_m|y_{1:t-1})$$

$$\sigma^2_{(SK)t-1} = E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\sigma)|y_{1:t-1}] = \sum_{m=1}^{M} \sigma^2_{t-1}(y_{1:t-2}, e'_m\theta_\sigma)q(\xi_{t-1} = e_m|y_{1:t-1}).$$

This approximation is similar to Dueker's approximation (Model D). As opposed to Dueker's approximation, we assume that σ^2_{t-1} is functions of $(y_{1:t-2}, \xi_{t-1})$. On the other hand, Klaassen's approximation reduces to the Simplified Klaassen approximation model by eliminating ξ'_t from the conditioning set.

2.4 Monte Carlo experiments on simulated data sets

We generate 100 independent time series of length $T = 500, 1500, 5000$ from the data generating process (DGP) corresponding to the model defined by equations (2.3) and (2.4) for two regimes ($M = 2$), time invariant transition probabilities and switching conditional mean and variance. In the simulation exercise, we set $\mu_t(y_{1:t-1}, \theta_\mu(s_t)) = (\mu_1, \mu_2) = (0.06, -0.09)$, $(\gamma_1, \gamma_2) = (0.30, 2.00)$, $(\alpha_1, \alpha_2) = (0.35, 0.10)$, $(\beta_1, \beta_2) = (0.20, 0.60)$, and $\pi_{11} = 0.98$, $\pi_{22} = 0.96$. This parameter setting corresponds to the one in Bauwens et al. [2010] for a similar Monte Carlo exercise. A relatively higher and more persistent conditional variance as compared to the first regime GARCH equation is implied by the second regime GARCH equation. Also, the probability of staying in each regime is close to one. For each of the simulated trajectories of the MS-GARCH model, the MLE estimates of the models in section 2.3 are obtained. Table 2.1 displays the summary statistics for 100 estimates of the MLE obtained for each of the auxiliary models.

Table 2.1: Mean and RMSE based on 100 estimates of the MLE.

T		DGP	Mean					RMSE				
		Values	Basic	Gray	Dueker	SimplifiedK	Klaassen	Basic	Gray	Dueker	SimplifiedK	Klaassen
500	π_{11}	0.98	0.9772	0.9772	0.9765	0.9768	0.9759	0.0121	0.0121	0.0133	0.0131	0.0153
	π_{22}	0.96	0.9482	0.9478	0.9507	0.9521	0.9498	0.0608	0.0612	0.0374	0.0282	0.0409
	μ_1	0.06	0.0637	0.0635	0.0621	0.0625	0.0541	0.0382	0.0380	0.0386	0.0389	0.0881
	μ_1	−0.09	−0.0770	−0.0773	−0.0869	−0.0920	−0.0795	0.2775	0.2824	0.2787	0.2764	0.2585
	γ_1	0.30	0.3338	0.3335	0.3359	0.3363	0.3072	0.0953	0.0950	0.0860	0.0878	0.0887
	γ_2	2.00	3.7611	3.7696	3.5854	3.6780	2.5465	2.4670	2.4789	2.2950	2.3927	1.6609
	α_1	0.35	0.3515	0.3514	0.3379	0.3330	0.3279	0.1429	0.1428	0.1441	0.1416	0.1381
	α_2	0.10	0.1090	0.1079	0.0933	0.0850	0.1090	0.1086	0.1085	0.1106	0.1090	0.1337
	β_1	0.20	0.0908	0.0912	0.1143	0.1080	0.1919	0.1555	0.1545	0.1566	0.1422	0.1646
	β_2	0.60	0.3579	0.3584	0.3765	0.3754	0.4977	0.3681	0.3677	0.3595	0.3736	0.3111
1500	π_{11}	0.98	0.9784	0.9783	0.9785	0.9787	0.9786	0.0065	0.0065	0.0065	0.0065	0.0065
	π_{22}	0.96	0.9544	0.9544	0.9548	0.9550	0.9561	0.0140	0.0140	0.0134	0.0133	0.0131
	μ_1	0.06	0.0600	0.0600	0.0601	0.0603	0.0596	0.0260	0.0260	0.0260	0.0259	0.0257
	μ_2	−0.09	−0.0918	−0.0916	−0.0972	−0.0983	−0.1041	0.1127	0.1126	0.1154	0.1165	0.1227
	γ_1	0.30	0.3407	0.3408	0.3430	0.3392	0.3006	0.0641	0.0641	0.0627	0.0614	0.0497
	γ_2	2.00	3.5961	3.5928	3.5505	3.5723	2.0785	1.8060	1.8038	1.7610	1.8096	0.8167
	α_1	0.35	0.3599	0.3599	0.3511	0.3472	0.3491	0.0707	0.0708	0.0716	0.0729	0.0717
	α_2	0.10	0.1149	0.1149	0.0971	0.0910	0.1016	0.0630	0.0630	0.0617	0.0621	0.0599
	β_1	0.20	0.0826	0.0824	0.0993	0.1041	0.1940	0.1271	0.1273	0.1151	0.1121	0.0852
	β_2	0.60	0.3858	0.3861	0.3967	0.4021	0.5836	0.2629	0.2627	0.2513	0.2559	0.1633
5000	π_{11}	0.98	0.9793	0.9793	0.9794	0.9795	0.9798	0.0030	0.0030	0.0030	0.0029	0.0030
	π_{22}	0.96	0.9575	0.9575	0.9579	0.9581	0.9588	0.0071	0.0071	0.0070	0.0069	0.0066
	μ_1	0.06	0.0611	0.0611	0.0609	0.0611	0.0603	0.0147	0.0147	0.0146	0.0147	0.0144
	μ_2	−0.09	−0.0800	−0.0799	−0.0817	−0.0828	−0.0888	0.0594	0.0594	0.0606	0.0604	0.0610
	γ_1	0.30	0.3507	0.3507	0.3514	0.3473	0.3081	0.0577	0.0577	0.0573	0.0539	0.0310
	γ_2	2.00	3.5882	3.5862	3.5109	3.5047	2.1107	1.6839	1.6820	1.6133	1.6240	0.5158
	α_1	0.35	0.3649	0.3649	0.3562	0.3525	0.3556	0.0417	0.0417	0.0398	0.0396	0.0389
	α_2	0.10	0.1204	0.1204	0.1027	0.0963	0.1036	0.0409	0.0409	0.0358	0.0368	0.0332
	β_1	0.20	0.0714	0.0713	0.0870	0.0920	0.1825	0.1320	0.1322	0.1178	0.1132	0.0550
	β_2	0.60	0.3837	0.3838	0.3992	0.4103	0.5796	0.2370	0.2368	0.2231	0.2186	0.0954

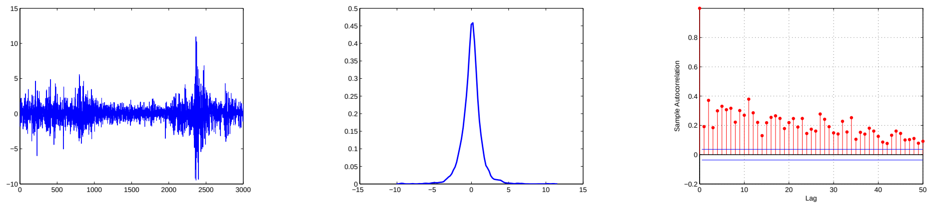
Notes: The estimated models belong to the class of MS-GARCH models described by equations 2.9 - 2.10(1). Results are based on 100 independent time series of length T=500, 1500 and 5000 from the DGP defined by equations 2.3 and 2.4. The parameter μ_i denotes the conditional mean, γ_i denotes the long run conditional variance in regime i, α_i and β_i are the regime specific ARCH and GARCH parameters, respectively, and the π_{ii} are the regime-staying probabilities.

In Table 2.1, we can observe that the estimates of γ_2 and β_2 across all models display significant biases. However, the biases decreases with increase in the sample size. This confirms the deduction by Augustyniak [2013] on a similar exercise that the MLE of GARCH models can be biased in small sample sizes. Furthermore, Table 2.1 shows that the Klaassens model is the most effective model among the set of models under consideration for generating consistent estimates for the path dependent MS-GARCH model. We further examine the performance of each these approximation by computing the percentage of correctly classified regime. An observation is assigned to regime 2 if its corresponding smoothed probability of being in that state is higher than the smoothed probability of staying in regime 1. We find that, on average, the auxiliary models correctly classified 91% of the observation for sample size $T = 500$ and 94% for sample size $T = 5000$. We also observe that there is no significant different in the number of correctly classified observation across the different auxilliary models. This observation may be argued to be a result of the low level of persistence in the GARCH models for each regime considered in this simulation exercise. To rule this out, we studied the results of parameterizations where one of regimes is highly persistent i.e. $\alpha+\beta$ is between 0.9 and 1. In line with this, 50 independent trajecttories of the MS-GARCH model with length $T = 5000$ was generated using the parameter set indicated in the third column of Table 2.2. A summary statistics for 50 estimates of the MLE obtained for each of the auxiliary models is reported in Table 2.2. Similar to our result in the low persistence case (Table 2.1), we observe, from Table 2.2, that Klaassen's model provides consistent estimates for a path dependent MS-GARCH model. We also observe that the percentage of correctly classified regimes by the auxiliary models are not significantly different from each other. i.e. the number vary between 94% and 96%.

Table 2.2: Mean and RMSE based on 50 estimates of the MLE.

T		DGP	Mean					RMSE				
		Values	Basic	Gray	Dueker	SimplifiedK	Klaassen	Basic	Gray	Dueker	SimplifiedK	Klaassen
5000	π_{11}	0.99	0.9892	0.9892	0.9894	0.9894	0.9891	0.0025	0.0025	0.0025	0.0025	0.0028
	π_{22}	0.99	0.9909	0.9909	0.9913	0.9913	0.9893	0.0029	0.0029	0.0030	0.0030	0.0028
	μ_1	0.06	0.0624	0.0625	0.0624	0.0624	0.0572	0.0193	0.0193	0.0192	0.0192	0.0346
	μ_2	−0.09	−0.0881	−0.0887	−0.0856	−0.0856	−0.0900	0.0596	0.0598	0.0582	0.0582	0.0698
	γ_1	0.40	0.6367	0.6367	0.6242	0.6225	0.4250	0.2444	0.2444	0.2315	0.2301	0.0782
	γ_2	0.60	2.4543	2.4487	2.2849	2.2674	0.6749	1.9412	1.9358	1.7773	1.7618	0.1531
	α_1	0.20	0.2158	0.2158	0.2071	0.2059	0.2010	0.0360	0.0360	0.0337	0.0334	0.0379
	α_2	0.05	0.0695	0.0695	0.0567	0.0544	0.0539	0.0259	0.0259	0.0178	0.0170	0.0288
	β_1	0.40	0.1091	0.1089	0.1303	0.1313	0.3851	0.2949	0.2951	0.2744	0.2735	0.1043
	β_2	0.90	0.7350	0.7354	0.7555	0.7626	0.8853	0.1770	0.1765	0.1575	0.1511	0.0555

Notes: The estimated models belong to the class of MS-GARCH models described by equations 2.9 - 2.10(1). Results are based on 50 replications, each consisting of 1500 observations from the DGP defined by equations 2.3 and 2.4. The parameter μ_i denotes the conditional mean, γ_i denotes the long run conditional variance in regime i, α_i and β_i are the regime specific ARCH and GARCH parameters, respectively, and the π_{ii} are the regime-staying probabilities.



(a) Time series (b) Density approximation (c) sample autocorrelation of squared returns

Figure 2.1: Graphs for S&P 500 daily returns from 20/05/1999 to 25/04/2011.

2.5 Empirical application: the S&P500 daily returns

In this section, Klaassen [2002] MS-GARCH model is fitted to S&P500 daily percentage returns from 20/05/1999 to 25/04/2011[1] (3000 observations). This daily data set has already been analyzed by several authors e.g. Bauwens et al. [2011], Dufays [2012] and Augustyniak [2013]. Thus, we compare the estimation results obtained with Klaassen [2002] MS-GARCH model to the MCEM-MCML algorithm developed by Augustyniak [2013]. Fig. 2.1 displays the returns sample path and kernel density estimate, and the autocorrelation of the squared returns. As one would expect from a typical financial time series, it exhibits strong persistence in the squared returns (see Fig. 2.1), slightly negative skewness and large excess kurtosis (see Tab. 2.3). These features calls for the use of a MS-GARCH model.

Table 2.3: Descriptive statistics for S&P 500 daily returns.

Min.	max.	Mean	Std.	Skewness	Kurtosis
−9.470	10.960	−0.00022	1.353	−0.116	10.546.

Fitting a full MS-GARCH model to empirical data can lead to parameters being estimated on the boundary of the parameter space and result in slow convergence. Based on this, a common practice in the estimation of MS-GARCH models with empirical data is to impose some restriction on the parameters. See Bauwens et al. [2010]

[1]Thanks to Bauwens et al. [2011] for making the data available on https://sites.google.com/site/websiteofarnauddufays/

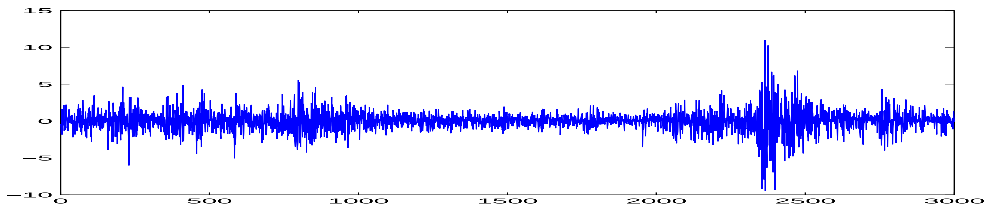
and Augustyniak [2013] for illustration. In this paper, two different specification of the MS-GARCH model are considered. The first is tagged unconstrained model. In this case, we provide estimates for all the parameters of the MS-GARCH model. The second corresponds to Augustyniak [2013] specification and we tag this constrained model. In this case we assume that both the ARCH and GARCH parameters are non-switching i.e. $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$.

Table 2.4: Estimation results of Klaassen [2002] model

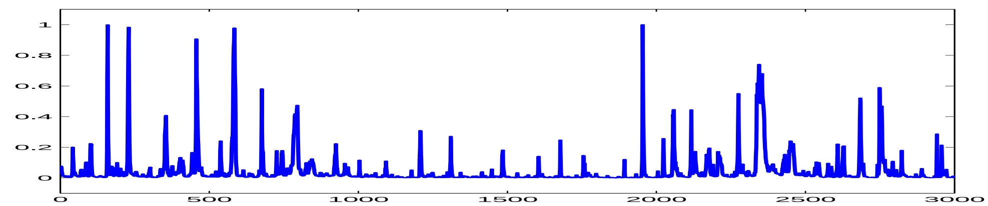
π_{11}	π_{22}	μ_1	μ_2	γ_1	γ_2	α_1	α_2	β_1	β_2	log-lik
Daily S&P 500: unconstrained model										
0.9738	0.8504	0.0660	−0.5319	0.0085	0.1864	0.0283	0.0560	0.9262	0.9440	−44487
(0.0122)	(0.0103)	(0.0006)	(0.0058)	(0.0091)	(0.0107)	(0.0190)	(0.0242)	(0.0082)	(0.0230)	
Daily S&P 500: constrained model with $\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$										
0.9894	0.7733	0.0421	−1.0092	0.0092	0.4142	0.0479	0.0479	0.9227	0.9227	−44558
(0.0067)	(0.0077)	(0.0003)	(0.0067)	(0.0053)	(0.0056)	(0.0053)	(0.0053)	(0.0036)	(0.0036)	

Notes: Standard errors in parentheses. "log-lik"denotes the log-likelihood of a model. The estimated model is based on Klaassen [2002] MS-GARCH model. The parameter μ_i denotes the conditional mean, γ_i denotes the long run conditional variance in regime i, α_i and β_i are the regime specific ARCH and GARCH parameters, respectively, and the π_{ii} are the regime-staying probabilities.

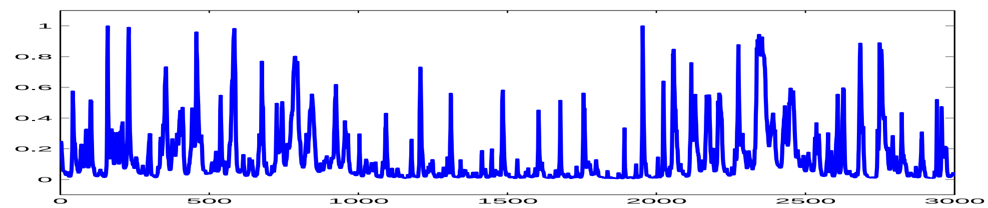
Table 2.4 reports the maximum likelihood estimates of both the constrained and unconstrained MS-GARCH model. The results displayed in Table 2.4 suggests that the estimation of the constrained model is very accurate, but more variability is observed for the unconstrained model. Furthermore, the results of our estimation on the constrained model are consistent with those obtained by Augustyniak [2013]. In both MS-GARCH specifications, we can identify the first regime as a positive mean returnlow volatility state while the second regime corresponds to a negative mean returnhigh volatility state. The second regime of the constrained model, however, exhibits (in absolute value) higher mean return (μ_2), higher long run unconditional variance (γ_2) and less persistence (i.e., π_{22} is reduced) when compared to the unconstrained model. The results of the log-likelihood ratio test suggests a better fit for



(a) Time series



(b) Constrained MS-GARCH model



(c) Unconstrained MS-GARCH model

Figure 2.2: MS-GARCH smoothed probabilities for being regime 2.

the unconstrained model. To illustrate the difference in the role played by regime two in the constrained and unconstrained MS-GARCH models, we can compare the smoothed inferences of the states given by these models (see Fig. 2.2). The trajectory of the smoothed probability of the models are quite similar. Nevertheless, the unconstrained model exhibit more noticeable and higher probability than the constrained model during some period suggesting that it is not as conservative as the constrained model. This shows that the unconstrained model is more flexible and gives more credible judgment about the possible state of this process.

2.6 conclusion

A family of MS-GARCH models based on collapsing procedure was developed. This class of MS-GARCH models encompasses some of the existing models in the literature and allows for the identification of new variants of the MS-GARCH model. Using simulated data, a comparison among the models belonging to this class was considered. The comparison is based on their ability to effectively and consistently generate estimates for a path dependent MS-GARCH model. It was shown that Klaassen's model provide the most accurate parameter estimates for a path dependent MS-GARCH model. However, we observed that all the auxiliary models produce equivalent number of correctly classified regime. This suggest that if one is interested in the value of the parameters then Klaassen's model is the choice model. On the other hand, if regime classification is of paramount interest then any of the auxiliary models can be used. Based on this observation, for empirical exercise we implement Klasssen's model.

Bibliography

A. Abramson and I. Cohen. On the stationarity of Markov-switching GARCH processes. *Econometric Theory*, 23:485–500, 2007.

D. Ardia. *Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications, volume 612 of Lecture Notes in Economics and Mathematical Systems.* Springer-Verlag, Berlin, Germany, 2008.

M. Augustyniak. Maximum likelihood estimation of the Markov-switching GARCH model. *Computational Statistics & Data Analysis*, 2013.

L. Bauwens, A. Preminger, and J. Rombouts. Theory and inference for a Markov switching GARCH model. *Econometrics Journal*, 13:218–244, 2010.

L. Bauwens, A. Dufays, and J. Rombouts. Marginal likelihood for Markov-switching and change-point GARCH. *CORE discussion paper, 2011/13*, 2011.

M. Bildirici and O. Ersin. Modeling markov switching arma-garch neural networks models and an application to forecasting stock returns. *Available at SSRN 2125855.*, 2012.

F. Black. The pricing of commodity contracts. *Journal of financial economics*, 3(1): 167–179, 1976.

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

T. Bollerslev and R. F.and D.B. Nelson Engle. *ARCH Models in Handbook of Econometrics, Vol.IV.* Amsterdam: North-Holland, 1994.

T. Bollerslev, R. Y. Chou, and K. F. Kroner. Arch modeling in finance: A review of the theory and empirical evidence. *Journal of Econometrics*, 52:5–59, 1992.

G. E. Box and D. R. Cox. An analysis of transformations (1964). an analysis of transformations. journal of the royal statistical society, 211-252. *Journal of the Royal Statistical Society*, Series B (Methodological):21152., 1964.

J. Cai. A Markov model of switching-regime ARCH. *Journal of Business and Economics Statistics*, 12:309–316, 1994.

M. Dueker. Markov switching in GARCH processes in mean reverting stock market volatility. *Journal of Business and Economics Statistics*, 15:26–34, 1997.

A. Dufays. Infinite-state Markov-switching for dynamic volatility and correlation models. *CORE discussion paper, 2012/43*, 2012.

R. J. Elliott, J. W. Lau, H. Miao, and T. K. Siu. Viterbi-based estimation for Markov switching GARCH models. *Applied Mathematical Finance*, 19(3):1–13, 2012.

R. F. Engle and T. Bollerslev. Modelling the persistence of conditional variances. *Econometric review*, 5(1):1–50, 1986.

R.F. Engle. Risk and volatility: Econometric models and financial practice. *American Economic Review*, 94:405–420, 2004.

C. Francq, M. Roussignol, and J. M. Zakoian. Conditional heteroskedasticity driven by hidden markov chains. *Journal of Time Series Analysis*, 22(2):197–220, 2001.

L. R. Glosten, R. Jagannathan, and D. E. Runkle. On the relation between the expected value and the volatility of the nominal excess return on stocks. *The Journal of Finance*, 48(5):1779–1801, 1993.

S. F. Gray. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42:27–62, 1996.

J. D. Hamilton and R. Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64:307–333, 1994.

C. He and T. Teräsvirta. Properties of moments of a family of garch processes. *Journal of Econometrics*, 92(1):173–192., 1999.

Z. He and J.M. Maheu. Real time detection of structural breaks in GARCH models. *Computational Statistics and Data Analysis*, 54(11):2628–2640, 2010.

F. Klaassen. Improving GARCH volatility forecasts with regime switching GARCH. *Empirical Economics*, 27:363–394, 2002.

C. Lamoureux and W. Lastrapes. Persistence in variance, structural change, and the GARCH model. *Journal of Business and Economics Statistics*, 8:225–234, 1990.

J. C. Liu. A family of markov?switching garch processes. *Journal of Time Series Analysis*, 33(6):892–902., 2012.

J. Marcucci. Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, 9(4):1558–3708, 2005.

T. Mikosch and C. Starica. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86: 378–390, 2004.

D. B. Nelson. Conditional heteroskedasticity in asset returns: A new approach. *Econometrica*, 59:347–370, 1991.

Chapter 3

Efficient Gibbs Sampling for Markov Switching GARCH Models

Abstract We develop efficient simulation techniques for Bayesian inference on switching GARCH models and contribute to the existing literature in several directions. First, we discuss different multi-move sampling techniques for Markov Switching (MS) state space models with particular attention to MS-GARCH models. Our multi-move sampling strategy is based on the Forward Filtering Backward Sampling (FFBS) applied to a auxiliary MS-GARCH model. Another important contribution is the use of multi-point samplers, such as the Multiple-Try Metropolis (MTM) and the Multiple-Trial Metropolized Independent Sampler, in combination with FFBS. In this sense we extend to MS state space models the work of So [2006] on efficient MTM sampler for continuous state space models. Finally, we suggest to further improve the sampler efficiency by introducing the antithetic sampling of Craiu and Meng [2005] and Craiu and Lemieux [2007] within the FFBS proposal. Our simulation experiments and application to financial data show the efficiency and effectiveness of our multi-point and multi-move strategies.

Keywords : Bayesian inference, GARCH, Markov switching, Multiple-try Metropolis

3.1 Introduction

The modeling of financial markets volatility has remained a prominent area of research in finance given the important role it plays in a variety of financial problems (e.g. asset pricing and risk management). Among volatility models, the Bollerslev [1986] Generalized Autoregressive Conditional Heteroskedastic (GARCH) model and its variants ranks as the most popular class of models among researchers and practitioners (e.g., see Ardia [2008]). However, from empirical studies, this class of models have been well documented to exhibit high persistence of conditional variance, i.e. the process is close to being non-stationary (nearly integrated). Lamoureux and Lastrapes [1990], among others, argue that the presence of structural changes in the variance process, for which the standard GARCH process cannot account for, may be responsible for this phenomenon. To buttress this point, Mikosch and Starica [2004] estimate a GARCH model on a sample that exhibits structural changes in its conditional variance and obtained a nearly integrated GARCH effect from the estimate. Based on this observation, Hamilton and Susmel [1994] and Cai [1994] propose a Markov Switching-Autoregressive Conditional Heteroskedastic (MS-ARCH) model, governed by a state variable that follows a first order Markov chain to capture the high volatility persistence, while Gray [1996] considers a Markov Switching GARCH (MS-GARCH) model since it can be written as an infinite order ARCH model and may be more parsimonious than the MS-ARCH model for financial data.

The class of MS-GARCH models is gradually becoming a work house among economics and financial practitioners for analysing financial markets data (e.g., see Marcucci [2005]). For practical implementation of this class of theoretical models, it is crucial to have reliable parameter estimators. Maximum Likelihood (ML) approach is a natural route to parameter estimation in Econometrics. However, the ML technique is not computationally feasible for MS-GARCH models because of the path dependence problem (see Gray [1996]). To this end, Henneke et al.

[2011] and Bauwens et al. [2010] propose Bayesian approach based on Markov Chain Monte Carlo (MCMC) Gibbs technique for estimating the parameters of Markov Switching-Autoregressive Moving Average-Generalized Autoregressive Conditional Heteroskedastic (MS-ARMA-GARCH) and MS-GARCH models respectively. Their proposed algorithm samples each state variable given others individually (single-move Gibbs sampler). This sampler is slowly converging and computationally demanding. Great attention have been paid in the literature at improving such inefficiencies in the context of continuous and possibly non-Gaussian and nonlinear state space models. See, for example, Frühwirth-Schnatter [1994], Koopman and Durbin [2000], De Jong and Shephard [1995] and Carter and Kohn [1994] for multi-move Gibbs sampler and So [2006] for multi-points and multi-move Gibbs sampling schemes for continuous and nonlinear state space models. To the best of our knowledge there are few works on efficient multi-move sampling scheme for discrete or mixed state space models. See Kim and Nelson [1999] for a review on multi-move Gibbs for conditionally linear models, Billio et al. [1999] for global Metropolis-Hastings algorithm for sampling the hidden states of MS-ARMA models and Fiorentini et al. [2012] for multi-move sampling in dynamic mixture models. As regards MS-GARCH models, Ardia [2008] develops a Gibbs sampling scheme for the joint sampling of the state variables for the Haas et al. [2004] model, which is a particular approximation of a MS-GARCH model, He and Maheu [2010] propose a Sequential Monte Carlo (SMC) algorithm for GARCH models subject to structural breaks, while Bauwens et al. [2011] propose a Particle MCMC (PMCMC) algorithm for estimating GARCH models subject to either structural breaks and regime switching. Dufays [2012], on the other hand, proposes a Metropolis-Hastings algorithm for block sampling of the hidden state of infinite state MS-GARCH models. See also Elliott et al. [2012] for an alternative approach, i.e. Viterbi-based technique, for sampling the state variables in MS-GARCH models.

In this paper, we develop an efficient simulation based estimation approach for MS-GARCH models with a finite number of regimes wherein the conditional mean and conditional variance of the observation process may change over time. We follow a data augmentation framework by including the state variables into the parameter vector. In particular, we propose a Bayesian approach based on MCMC algorithm which allows us to circumvent the problem of path dependence by simultaneously generating the states (multi-move Gibbs sampler) from their joint distribution. Our strategy for sampling the state variables is based on Forward Filtering Backward Sampling (FFBS) techniques. As in the case of mixed hidden state models, FFBS algorithm cannot be applied directly on switching GARCH models, thus we suggest the use of a Metropolis algorithm with an FFBS proposal generated using an auxiliary model. We propose and discuss different auxiliary models obtained by alternative approximations of the MS-GARCH conditional variance equation.

Another original contribution of the paper relates to the Metropolis step for the hidden states. To efficiently estimate MS-GARCH models we consider the class of generalized (multipoint) Metropolis algorithms (see Liu [2002], Chapter 5) which extends the standard Metropolis-Hastings (MH) approach (Hastings [1970] and Metropolis et al. [1953]). See Liu [2002] and Robert and Casella [2007] for an introduction to MH algorithms and a review of various extensions. Multipoint samplers have been proved, both theoretically and computationally, to be effective in improving the mixing rate of the MH chain and the efficiency of the Monte Carlo estimates based on the output of the chain. The main feature of the multipoint samplers is that at each iteration of the MCMC chain the new value of the chain is selected among multiple proposals, while in the MH algorithm one accepts or rejects a single proposal. In this paper we apply the Multiple-Try Metropolis (MTM) (see [Liu et al., 2000]) and some modified MTM algorithms. The superiority of the MTM over standard MH algorithm has been proved in Craiu and Lemieux [2007], who also propose to apply antithetic and quasi-Monte Carlo techniques to obtain good proposals in the MTM. So [2006] applies MTM to the estimation of latent-variable models and finds evidence of superiority of the MTM over standard MH samplers for the latent variable estimation. The author also finds that the efficiency of MTM can further be increased by the use of multi-move sampling. Casarin et al. [2013] apply the MTM transition to the context of interacting chains. They provide a comparison with standard interacting MH and also estimate the gain of efficiency when using interacting MTM combined with block-sampling for the estimation of stochastic volatility models. We thus combine the MTM sampling strategies with the approximated FFBS techniques for the Markov switching process. In this sense, we extend the work of So [2006] to the more complex case of Markov-switching nonlinear state space models. In fact, the use of multiple proposals is particularly suited in this context where the forward filter is used at each iteration to generate only one proposal with a large computational cost. The use of multiple proposals based on the same run of the forward filter is thus discussed. We also apply to this context the antithetic sampling technique proposed by Craiu and Lemieux [2007] to generate negatively correlated proposal within the Multiple-try algorithm, and suggest a Forward Filtering Backward Antithetic Sampling (FFBAS) algorithm which combines the permuted displacement algorithm of Craiu and Meng [2005] with FFBS and possibly produces pairwise negative association among the trajectories of the hidden states. Note that our approach could easily be extended to other discrete or mixed state space models. Our efficient sampling procedure may also be applied to simulation-based maximum likelihood context (such as Billio et al. [1998] and Augustyniak [2013]) to carry out inference on MS-GARCH model.

The paper is organized as follows. Section 3.2 introduces the MS-GARCH model and discuss inference issues related to existing methods in the literature. In Section 3.3, we present the Bayesian inference approach and explain the multi-move multi-point sampling strategies. In Section 3.4, we study the efficiency of our estimation procedure through some simulation experiments. In Section 3.5 we provide an application to S&P 500 return series while in Section 3.6, we conclude and discuss possible extensions.

3.2 Markov switching GARCH models

3.2.1 The model

A Markov Switching GARCH model is a nonlinear specification of the evolution of a time series assessed to be affected by different states of the world and for which the conditional variance in each state follows a GARCH process. More specifically, let y_t be the observed variable (e.g. the return on some financial asset) and s_t a discrete, unobserved, state variable which could be interpreted as the state of the world at time t. Define $(y_t, \ldots, y_u)$ and $(s_t, \ldots, s_u)$ as $y_{t:u}$ and $s_{t:u}$ respectively whenever $t \leq u$. Then

$$y_t = \mu_t(y_{1:t-1}, \theta_\mu(s_t)) + \sigma_t(y_{1:t-1}, \theta_\sigma(s_t))\eta_t, \qquad \eta_t \overset{iid}{\sim} N(0,1), \tag{3.1}$$

$$\sigma_t^2(y_{1:t-1}, \theta_\sigma(s_t)) = \gamma(s_t) + \alpha(s_t)\epsilon_{t-1}^2 + \beta(s_t)\sigma_{t-1}^2(y_{1:t-2}, \theta_\sigma(s_{t-1})), \tag{3.2}$$

where, $N(0,1)$ denotes the standard normal distribution, θ_μ represents the parameters of the conditional mean of the returns series, $\epsilon_t = \sigma_t(y_{1:t-1}, \theta_\sigma(s_t))\eta_t$, $\theta_\sigma(s_t) = (\gamma(s_t), \alpha(s_t), \beta(s_t))$, $\gamma(s_t) > 0$, $\alpha(s_t) \geq 0$, $\beta(s_t) \geq 0$, and $s_t \in \{1, \ldots, M\}$, $t = 1, \ldots, T$, is assumed to follow a M-state first order Markov chain with transition probabilities $\{\pi_{ij,t}\}_{i,j=1,2,\ldots,M}$:

$$\pi_{ij,t} = p(s_t = i | s_{t-1} = j, y_{1:t-1}, \theta_\pi), \quad \sum_{i=1}^{M} \pi_{ij,t} = 1 \quad \forall\ j = 1, 2, \ldots, M,$$

where θ_π represents the parameters of the transition matrix. The parameter shift functions $\gamma(s_t)$, $\alpha(s_t)$ and $\beta(s_t)$, describe the dependence of parameters on the regime s_t i.e.

$$\gamma(s_t) = \sum_{m=1}^{M} \gamma_m \mathbb{I}_{s_t=m}, \quad \alpha(s_t) = \sum_{m=1}^{M} \alpha_m \mathbb{I}_{s_t=m}, \quad \text{and} \quad \beta(s_t) = \sum_{m=1}^{M} \beta_m \mathbb{I}_{s_t=m},$$

where,

$$\mathbb{I}_{s_t=m} = \begin{cases} 1, & \text{if } s_t = m, \\ 0, & \text{otherwise.} \end{cases}$$

By defining the allocation variable, s_t, as a M-dimensional discrete vector, $\xi_t = (\xi_{1t}, \ldots, \xi_{Mt})'$, where $\xi_{mt} = \mathbb{I}_{s_t=m}$, $m = 1, \ldots, M$, the system of equations in (4.2)-(4.6) can be written compactly as

$$y_t = \mu_t(y_{1:t-1}, \xi_t'\theta_\mu) + \sigma_t(y_{1:t-1}, \xi_t'\theta_\sigma)\eta_t, \qquad \eta_t \overset{iid}{\sim} N(0,1), \tag{3.3}$$

$$\sigma_t^2(y_{1:t-1}, \xi_t'\theta_\sigma) = (\xi_t'\gamma) + (\xi_t'\alpha)\epsilon_{t-1}^2 + (\xi_t'\beta)\sigma_{t-1}^2(y_{1:t-2}, \xi_{t-1}'\theta_\sigma), \tag{3.4}$$

where $\epsilon_t = \sigma_t(y_{1:t-1}, \xi_t'\theta_\sigma)\eta_t$, $\gamma = (\gamma_1, \ldots, \gamma_M)'$, $\alpha = (\alpha_1, \ldots, \alpha_M)'$, $\beta = (\beta_1, \ldots, \beta_M)'$, $\theta_\mu = (\theta_{1\mu}, \ldots, \theta_{M\mu})'$ and $\theta_\sigma = (\theta_{1\sigma}, \ldots, \theta_{M\sigma})'$ with $\theta_{m\sigma} = (\gamma_m, \alpha_m, \beta_m)'$ for $m = 1, \ldots, M$. for $t = 1, \ldots, T$. Let $\pi_t = (\pi_{1t}, \ldots, \pi_{Mt})$, with $\pi_{it} = (\pi_{i1,t}, \ldots, \pi_{iM,t})$ for $i = 1, 2, \ldots, M$ and $\sum_{i=1}^{M} \pi_{ij,t} = 1$ for all $j = 1, 2, \ldots, M$. Since ξ_t follows a $M-$state first order Markov chain, we define the transition probabilities $\{\pi_{ij,t}\}_{i,j=1,2,\ldots,M}$ by

$$\pi_{ij,t} = p(\xi_t = e_i | \xi_{t-1} = e_j, y_{1:t-1}, \theta_\pi),$$

where e_i is the $i-$th column of a M-by-M identity matrix. The conditional probability of ξ_t given ξ_{t-1}, θ_π and $y_{1:t-1}$ is given by

$$p(\xi_t|\xi_{t-1}, y_{1:t-1}, \theta_\pi) = \prod_{m=1}^{M} (\pi_{mt}\xi_{t-1})^{\xi_{mt}}, \tag{3.5}$$

which implies that the probability with which event m occurs at time t is $\pi_{mt}\xi_{t-1}$.

3.2.2 Inference issues

Estimating Markov switching GARCH models is a challenging problem since the likelihood of y_t depends on the entire sequence of past states up to time t due to the recursive structure of its volatility. To elaborate on this, the likelihood function of the switching GARCH model is given by

$$\mathcal{L}(\theta|y_{1:T}) \equiv f(y_{1:T}|\theta) = \sum_{i_1=1}^{M} \cdots \sum_{i_T=1}^{M} f(y_{1:T}, \xi_1 = e_{i_1}, \ldots, \xi_T = e_{i_T}|\theta) \tag{3.6}$$

where $\theta = (\{\theta_{m\mu}, \theta_{m\sigma}\}_{m=1,\ldots,M}, \theta_\pi)$. Setting $\xi_{s:t} = (\xi_s', \ldots, \xi_t')$ whenever $s \leq t$, the joint density function of $y_{1:t}$ and $\xi_{1:t}$ on the right hand side of equation (3.6) is

$$\begin{aligned} f(y_{1:T}, \xi_{1:T}|\theta) &= f(y_1|\xi_{1:1}, \theta_\mu, \theta_\sigma) \prod_{t=2}^{T} f(y_t|y_{1:t-1}, \xi_{1:t}, \theta_\mu, \theta_\sigma) p(\xi_t|y_{1:t-1}, \xi_{1:t-1}, \theta_\pi) \\ &= f(y_1|\xi_{1:1}, \theta_\mu, \theta_\sigma) \prod_{t=2}^{T} f(y_t|y_{1:t-1}, \xi_{1:t}, \theta_\mu, \theta_\sigma) \left(\prod_{i=1}^{M} (\pi_{it}\xi_{t-1})^{\xi_{it}} \right), \end{aligned} \tag{3.7}$$

with,

$$f(y_t|y_{1:t-1}, \xi_{1:t}, \theta_\mu, \theta_\sigma) \propto \frac{1}{\sigma_t(y_{1:t-1}, \xi_t'\theta_\sigma)} \exp\left(-\frac{1}{2} \left(\frac{y_t - \mu_t(y_{1:t-1}, \xi_t'\theta_\mu)}{\sigma_t(y_{1:t-1}, \xi_t'\theta_\sigma)} \right)^2 \right).$$

Given σ_1, recursive substitution in equation (3.4) yields

$$\sigma_t^2 = \sum_{i=0}^{t-2} \left[\xi'_{t-i}\gamma + (\xi'_{t-i}\alpha)\epsilon^2_{t-1-i}\right] \prod_{j=0}^{i-1} \xi'_{t-j}\beta + \sigma_1^2 \prod_{i=0}^{t-2} \xi'_{t-i}\beta. \tag{3.8}$$

Equation (3.8) clearly shows the dependence of conditional variance at time t on the entire history of the regimes and thus the dependence of the likelihood function on the entire history of the regimes. The evaluation of the likelihood function over a sample of length T, as can be seen in equation (3.6), involves integration (summation) over all M^T unobserved states i.e. integration over all M^T possible (unobserved) regime paths. This requirement makes the maximum likelihood estimation of θ infeasible in practice.

Two major approaches have been developed in the literature in order to circumvent this path dependence problem. One approach involves the use of model approximation while the other is simulation based.

As regards to the model approximation approach, Cai [1994] and Hamilton and Susmel [1994] approximated the MS-GARCH model by an MS-ARCH model. This approach effectively makes the model tractable because the lagged conditional variance that makes the conditional variance dependent on the history of regime has been dropped. Kaufman and Frühwirth-Schnatter [2002] employed the algorithm developed by Chib [1996] for a Markov mixture models to compute the marginal likelihood of the MS-ARCH model but noted that this methodology cannot be carried over to the MS-GARCH model because of the path dependence problem. Another approximation approach can be credited to Gray [1996] who noted that the conditional density of the return is essentially a mixture of distributions with time-varying mixing parameter and in particular under normality assumption he suggested the use of aggregate conditional variances over all regimes as the lagged conditional variance when constructing the conditional variance at each time step. Extensions of Gray

[1996] model can be found in Dueker [1997], Klaassen [2002] and Haas et al. [2004] among others. Abramson and Cohen [2007] provide stationarity conditions for some of these approximations. The problem with this approach is that these approximations cannot be verified.

Among the simulation based approaches proposed in the literature there is the Bayesian estimation technique proposed by Bauwens et al. [2010]. In particular, they develop a single-move MCMC Gibbs sampler for a Markov switching GARCH model with a fixed number of regimes. The authors also provide sufficient conditions for geometric ergodicity and existence of moments of the process. Their estimation approach, though quite promising, has one main limitation that has rendered it unattractive. The single-move Gibbs sampler is inefficient i.e. draws from the single-move scheme are noted to be highly correlated and thus slow down the convergence of the Markov chain. Alternative simulation based approaches relies on particle filter also known as sequential Monte Carlo methods. He and Maheu [2010] develop a sequential Monte Carlo method for estimating GARCH models subject to an unknown number of structural breaks. Bauwens et al. [2011] propose a particle MCMC approach to GARCH models subject to structural break or regime switching. Augustyniak [2013] developed a hybrid Monte Carlo expectation-maximization and Monte Carlo maximum likelihood algorithm to calculate the maximum likelihood estimator the MS-GARCH model.

3.3 Bayesian inference

Based on the aforementioned inference issues associated with MS-GARCH models, we present a Bayesian approach based on MCMC Gibbs algorithm which allows us to circumvent the path dependence problem and efficiently sample the state trajectory. The purpose of this algorithm is to generate samples from the posterior distribution which are then used for its characterization. We follow a data augmentation framework by treating the state variables as parameters of the model and construct the likelihood function assuming the states known.

Before proceeding with the elicitation of our proposed Bayesian technique, it is important that we make explicit the parametric specification of the conditional mean, $\mu_t(y_{1:t-1}, \xi_t'\theta_\mu)$, of the return process y_t in equation (3.3) and the transition probabilities $p(\xi_t'|\xi_{t-1}', y_{1:t-1}, \theta_\pi)$. Since our major aim is to define a technique for sampling the state variables efficiently, which in turn will affect other parameter estimates, we assume for expository purposes a conditional mean defined by a constant switching parameter given by $\xi_t'\mu$ where $\mu = (\mu_1, \ldots, \mu_M)'$ and constant transition probabilities. Alternative specification such as switching ARMA process could be thought of for the conditional mean and time varying transition probabilities may be defined by following Gray [1996] approach, i.e. specifying transition probabilities as a function of past observables. Under this specification, the augmented parameter set of our model consists of $\xi_{1:T}$, $\theta = (\theta_\mu, \theta_\sigma, \theta_\pi)$ where $\theta_\mu = \mu$, $\theta_\pi = (\{\pi_m\}_{m=1,\ldots,M})$ and $\theta_\sigma = (\{\theta_{m\sigma}\}_{m=1,\ldots,M})$ with $\theta_{m\sigma} = (\gamma_m, \alpha_m, \beta_m)$, $\pi_m = (\pi_{m1}, \ldots, \pi_{mM})$ and $\sum_{m=1}^{M} \pi_{mm^*} = 1 \; \forall \; m^* = 1, \ldots, M$. We assume a fairly informative prior distributions for the transition probabilities θ_π

$$\theta_\pi \sim \prod_{m=1}^{M} \text{Dirichlet}(\nu_{1m}, \ldots, \nu_{Mm}),$$

where $\nu_{1m}, \ldots, \nu_{Mm} \; \forall \; m = 1, \ldots, M$ are hyperparameters to be defined. Other parameters of the MS-GARCH model are assumed to follow independent uniform priors. If the uniform is defined on a bounded domain (see Bauwens et al. [2010]) then the posterior is alway defined. If the uniform is an improper prior the posteriors are not always proper distributions (see Wasserman [2000]). The posterior of the k-thregime parameter is not proper if there are no observations allocated to the k-th regime. It is possible to avoid this offensive grouping of the data by rejecting, at each iteration of the Gibbs sampler, the draws of the sequence of allocation variables, ξ_t, $t = 1, \ldots, T$, that do not belong to the set $\mathcal{S} = \{\xi_{1:T} | \sum_{t=1}^{T} \xi_{jt} \geq 1, \forall j = 1, \ldots, M\}$. Moreover, to avoid the label switching problem, we assume this $\gamma_1 < \ldots < \gamma_M$ identifiability restriction. See Frühwirth-Schnatter [2006] for an introduction to label switching problem for dynamic mixtures and MS models and Bauwens et al. [2010] for illustration of the identification constraint for MS-GARCH models. The joint prior distribution is thus proportional to

$$f(\theta) \propto \prod_{m=1}^{M} \text{Dirichlet}(\nu_{1m}, \ldots, \nu_{Mm}) \mathbb{I}_{\gamma_1 < \ldots < \gamma_M}, \tag{3.9}$$

and the posterior density of the augmented parameter vector given by

$$f(\theta, \xi_{1:T} | y_{1:T}) \propto f(y_{1:T} | \xi_{1:T}, \theta) f(\xi_{1:T} | \theta) f(\theta), \tag{3.10}$$

cannot be identified with any standard distribution, hence we cannot sample directly from it. Using the Gibbs sampler, we can generate samples from this high-dimensional posterior density. This will be done by iteratively sampling from the following three full conditional distributions

1. $p(\xi_{1:T} | \theta, y_{1:T})$,

2. $f(\theta_\pi | \theta_\mu, \theta_\sigma, \xi_{1:T}, y_{1:T}) = f(\theta_\pi | \xi_{1:T})$,

3. $f(\theta_\sigma, \theta_\mu | \theta_\pi, \xi_{1:T}, y_{1:T}) = f(\theta_\sigma, \theta_\mu | \xi_{1:T}, y_{1:T})$.

These full conditional distributions are easier to manage and sample from because they can either be associated with a known distribution or simulated by a lower dimensional auxiliary sampler. In the following subsections we present in details our sampling procedure for both the single move and multi move algorithms .

3.3.1 Sampling the state variables $\xi_{1:T}$.

To sample $\xi_{1:T}$ using the single-move algorithm, one relies on computing

$$p(\xi_t|\xi_{1:t-1}, \xi_{t+1:T}, \theta, y_{1:T}) \propto \prod_{m=1}^{M} (\pi_m \xi_{t-1})^{\xi_{mt}} (\pi_m \xi_t)^{\xi_{m,t+1}} \prod_{j=t}^{T} f(y_j|\xi_j, \theta, y_{1:j-1}), \quad (3.11)$$

for each value ξ_t in $\{e_m : m = 1, \ldots, M\}$ and dividing each evaluation by the sum of the M points to get the normalized discrete distribution of ξ_t from which to sample. Sampling from such a distribution once the probabilities are known is similar to sampling from a Multinomial distribution. On the other hand, the full joint conditional distribution of the state variables, $\xi_{1:T}$, given the parameter values and return series

$$p(\xi_{1:T}|\theta, y_{1:T}) \propto f(y_{1:T}|\xi_{1:T}, \theta) p(\xi_{1:T}|\theta) \quad (3.12)$$

is a non-standard distribution. Therefore multi-move FFBS sampling is not feasible. For this reason, we consider a generalized metropolis (i.e. multipoint Metropolis) algorithm for generating the state variables. Multipoint samplers are designed to consider multiple proposals at each iteration of a Metropolis and to choose the new value of the chain from this trial set. The multi-move and multipoint sampling procedures are of interest because of their potentials at addressing issues associated with multi-modality of the target function (i.e. in the event that the target distribution is multi-modal in nature the MCMC chain runs the risk of getting trapped in local modes) and autocorrelation of samples from the Metropolis chain. Our scheme generally involves running a FFBS on the auxiliary sampler to generate several proposals at each iteration step. Let the proposal distribution be denoted by

$$q(\xi_{1:T}|\theta, y_{1:T}) = q(\xi_T|\theta, y_{1:T}) \prod_{t=1}^{T-1} q(\xi_t|\xi_{t+1}, \theta, y_{1:t}), \quad (3.13)$$

where $q(\xi_t|\xi_{t+1}, \theta, y_{1:t}) \propto q(\xi_t|y_{1:t}, \theta)q(\xi_{t+1}|\xi_t, \theta)$ with $q(\xi_t'|y_{1:t}, \theta)$ representing filtered probability. A discussion on the proposal distribution is presented in Section 3.3.2. In the following, we discuss the three multipoint algorithms considered in this paper.

Multiple-try Metropolis sampler

Liu et al. [2000] suggest the Multiple-Try Metropolis (MTM) sampler scheme. As in the general case of multipoint samplers, their idea is to consider several points generated by a proposal distribution so that possibly a larger region from which the new value for the chain is chosen can be investigated. By using the multiple-try strategy, it is easier for the iterates to jump from one local maximum to another and thus speed up the convergence to the desired target distribution. Samples from the proposal distribution will be generated by FFBS algorithm. We present below a sketch of the main ingredients needed in Forward Filter (FF) and Backward Sampling (BS) algorithm and refer the reader to Frühwirth-Schnatter [2006] for detailed presentation of this procedure. At time t, given θ and $y_{1:t}$ the FF probabilities are obtained by first computing the one-step ahead prediction

$$q(\xi_t|\theta, y_{1:t-1}) = \sum_{i=1}^{M} \left(\prod_{j=1}^{M} (\pi_j e_i)^{\xi_{j,t}} \right) q(\xi_{t-1} = e_i|\theta, y_{1:t-1}),$$

then, the FF is

$$q(\xi_t|\theta, y_{1:t}) = \frac{g(y_t|\xi_t, \theta, y_{1:t-1})q(\xi_t|\theta, y_{1:t-1})}{\sum_{i=1}^{M} g(y_t|\xi_t = e_i, \theta, y_{1:t-1})q(\xi_t = e_i|\theta, y_{1:t-1})}, \tag{3.14}$$

where $g(y_t|\xi_t, \theta, y_{1:t-1})$ is the conditional density of the return process under the auxiliary model. Using the output of the FF, we compute $q(\xi_T|\theta, y_{1:T})$ and

$$q(\xi_t|\xi_{t+1}, \theta, y_{1:t}) = \frac{\prod_{j=1}^{M} (\pi_j \xi_t)^{\xi_{j,t+1}} q(\xi_t|\theta, y_{1:t})}{q(\xi_{t+1}|\theta, y_{1:t})}, \tag{3.15}$$

for $t = T-1, T-2, \ldots, 2, 1$. Then at each time step we sample ξ_T from $q(\xi_T|\theta, y_{1:T})$ and ξ_t from $q(\xi_t|\xi_{t+1}, \theta, y_{1:t})$ iteratively for $t = T-1, T-2, \ldots, 2, 1$. This is the BS step. The BS procedure is implemented by first noting that ξ_{t+1} is the most recent value sampled for the hidden Markov chain at $t+1$ and since ξ_t can take one of $e_1, \ldots, e_M$, we compute the expression in equation (3.15) for each of these values. Sampling ξ_t from $q(\xi_t|\xi_{t+1}, \theta, y_{1:t})$ may be compared to multinomial sampling, provided that the probability of $\xi_i = e_i$, $i = 1, \ldots, M$, are known. Note that at each iteration step of the MCMC procedure we only need a single run of the Forward Filter (FF) for generating multiple proposals using Backward Sampling (BS).

A summary of our MTM algorithm is given in Algorithm 1.

Algorithm 1 Multiple Try Metropolis Sampler (MTM)

i. Choose a starting value $\xi^0_{1:T}$.

ii. Let $\xi^{(r-1)}_{1:T}$ be the value of the MTM at the $(r-1)$-th iteration.

iii. Construct a trial set $\{\xi_{1:T,1}, \xi_{1:T,2}, \ldots, \xi_{1:T,K}\}$ containing K independent paths of the state variable drawn from the proposal distribution $q(\xi_{1:T}|\theta^{(r-1)}, y_{1:T})$.

iv. Evaluate

$$W_k(\xi_{1:T,k}, \xi^{(r-1)}_{1:T}) = \frac{p(\xi_{1:T,k}|\theta^{(r-1)}, y_{1:T})}{q(\xi_{1:T,k}|\theta^{(r-1)}, y_{1:T})}, \quad \forall k = 1, \ldots, K.$$

v. Select $\tilde{\xi}_{1:T}$ from $\{\xi_{1:T,1}, \xi_{1:T,2}, \ldots, \xi_{1:T,K}\}$ according to the probability

$$p_k = \frac{W_k(\xi_{1:T,k}, \xi^{(r-1)}_{1:T})}{\sum_{k=1}^K W_k(\xi_{1:T,k}, \xi^{(r-1)}_{1:T})}, \quad \forall k = 1, \ldots, K.$$

vi. Construct a reference set $\{\xi^*_{1:T,1}, \xi^*_{1:T,2}, \ldots, \xi^*_{1:T,K}\}$ by setting the first $K-1$ elements to a new set of samples drawn from the proposal distribution $q(\xi_{1:T}|\theta^{(r-1)}, y_{1:T})$ and the $K-$th element $\xi^*_{1:T,K}$ to $\xi^{(r-1)}_{1:T}$.

vii. Draw $u \sim \mathcal{U}_{[0,1]}$ and set

$$\xi^{(r)}_{1:T} = \begin{cases} \tilde{\xi}_{1:T} & \text{if } \; u \leq \alpha(\tilde{\xi}_{1:T}, \xi^{(r-1)}_{1:T}), \\ \xi^{(r-1)}_{1:T} & \text{otherwise,} \end{cases}$$

where,

$$\alpha(\tilde{\xi}_{1:T}, \xi^{(r-1)}_{1:T}) = \min\left(1, \frac{\sum_{k=1}^K W_k(\xi_{1:T,k}, \xi^{(r-1)}_{1:T})}{\sum_{k=1}^K W_k(\xi^*_{1:T,k}, \tilde{\xi}_{1:T})}\right).$$

Observe that the MTM algorithm reduces to standard Metropolis-Hasting algorithm when $K = 1$. We also note that alternative weight function other than the importance weight function assumed in the MTM algorithm presented above could be used (e.g., see Craiu and Lemieux [2007]).

Multiple-trial Metropolized independent sampler

As suggested by Liu [2002], when using independent proposal distributions the generation of a set of reference points is not needed to have a possibly more efficient generalized Metropolis algorithm. Thus, we combine the FFBS proposals with the Liu [2002] Metropolized Independent Sampler (MTMIS) and obtain Algorithm 2. The main advantage is that one can use multiple proposals without generating the reference points, obtaining thus a decrease of the computational complexity of the algorithm.

Algorithm 2 Multiple-Trial Metropolized Independent Sampler (MTMIS)

i. Choose a starting value $\xi^0_{1:T}$.

ii. Let $\xi^{(r-1)}_{1:T}$ be the value of the MTM at the $(r-1)$-th iteration.

iii. Construct a trial set $\{\xi_{1:T,1}, \xi_{1:T,2}, \dots, \xi_{1:T,K}\}$ containing K independent paths of the state variable drawn from the proposal distribution $q(\xi_{1:T}|\theta^{(r-1)}, y_{1:T})$.

iv. Evaluate

$$W_k(\xi_{1:T,k}) = \frac{p(\xi_{1:T,k}|, \theta^{(r-1)}, y_{1:T})}{q(\xi_{1:T,k}|\theta^{(r-1)}, y_{1:T})}, \quad \forall \;\; k = 1, \dots, K,$$

and define

$$W = \sum_{k=1}^{K} W_k(\xi_{1:T,k})$$

v. Select $\tilde{\xi}_{1:T}$ from $\{\xi_{1:T,1}, \xi_{1:T,2}, \dots, \xi_{1:T,K}\}$ according to the probability

$$p_k = \frac{W_k(\xi_{1:T,k})}{\sum_{k=1}^{K} W_k(\xi_{1:T,k})}, \quad \forall k = 1, \dots, K.$$

vi. Draw $u \sim \mathcal{U}_{[0,1]}$ and set

$$\xi^{(r)}_{1:T} = \begin{cases} \tilde{\xi}_{1:T} & \text{if } \; u \leq \alpha(\tilde{\xi}_{1:T}, \xi^{(r-1)}_{1:T}) \\ \xi^{(r-1)}_{1:T} & \text{otherwise} \end{cases}$$

where,

$$\alpha(\tilde{\xi}_{1:T}, \xi^{(r-1)}_{1:T}) = \min\left(1, \frac{W}{W - W(\tilde{\xi}_{1:T}) + W(\xi^{(r-1)}_{1:T})}\right).$$

Multiple correlated-try Metropolis sampler

To further improve the efficiency the MTM algorithm and to ensure that a larger portion of the sample space is explored for better mixing and shorter running time, we propose the use of correlated proposals. There are various ways of introducing correlation among proposals e.g. antithetic and stratified approaches. In this paper, we study the antithetic approach. The use of antithetic sampling in a Gibbs sampling context allows for a gain of efficiency. Pitt and Shephard [1996] propose a blocking method with antithetic approach for non-Gaussian state space models, Holmes and Jasra [2009] propose a scheme for reducing the variance of estimates from the standard Metropolis-within-Gibbs sampler by introducing antithetic samples while Bizjajeva and Olsson [2008] propose a forward filtering backward smoothing particle filter algorithm with antithetic proposal. Here we follow Craiu and Lemieux [2007] which use antithetic proposals within a multi-point sampler and apply their idea to the context of discrete state space models. We propose Multiple Correlated-Try Metropolis sampler (MCTM) based on a combination of the FFBS and antithetic sampling techniques. To the best of our knowledge, antithetic proposals of this kind have not been used in the context of Markov switching nonlinear state space models. The idea is to choose, at each step of the MCMC algorithm, a new hidden state trajectory from negatively correlated proposals instead of independent proposals. Following the suggestion of Liu [2002], we obtain Algorithm 3.

Algorithm 3 Multiple Correlated-Try Metropolis Sampler (MCTM)

i. Choose a starting value $\xi^0_{1:T}$.

ii. Let $\xi^{(r-1)}_{1:T}$ be the value of the MTM at the $(r-1)$-th iteration.

iii. Construct a trial set $\{\xi_{1:T,1}, \xi_{1:T,2}, \dots, \xi_{1:T,K}\}$ containing K correlated paths of the state variable drawn from the proposal distribution.

iv. Evaluate

$$W_1(\xi_{1:T,1}) = p(\xi_{1:T,1}|\theta^{(r-1)}, y_{1:T}),$$
$$W_k(\xi_{1:T,1:k}) = p(\xi_{1:T,k}|\theta^{(r-1)}, y_{1:T}) \prod_{i=1}^{k-1} q(\xi_{1:T,i}|\theta^{(r-1)}, y_{1:T}, \xi_{1:T,i:k}) \;\; \forall \;\; k = 2, \dots, K,$$

v. Select $\tilde{\xi}_{1:T}$ from $\{\xi_{1:T,1}, \xi_{1:T,2}, \dots, \xi_{1:T,K}\}$ according to the probability

$$p_k = \frac{W_k(\xi_{1:T,1:k}, \xi^{(r-1)}_{1:T})}{\sum_{k=1}^{K} W_k(\xi_{1:T,1:k}, \xi^{(r-1)}_{1:T})}, \quad \forall k = 1, \dots, K.$$

vi. Suppose $\tilde{\xi}_{1:T} = \xi_{1:T,l}$ is chosen in item (v) above, create a reference set $\{\xi^*_{1:T,1}, \xi^*_{1:T,2}, \dots, \xi^*_{1:T,K}\}$ by letting

$$\xi^*_{1:T,j} = \xi_{1:T,l-1} \quad \forall \;\; j = 1, \dots, l-1$$
$$\xi^*_{1:T,l} = \xi^{(r-1)}_{1:T}$$

and drawing $\xi^*_{1:T,j}$ for $j = l+1, \dots, K$ from the proposal distribution.

vii. Draw $u \sim \mathcal{U}_{[0,1]}$ and set

$$\xi^{(r)}_{1:T} = \begin{cases} \tilde{\xi}_{1:T} & \text{if } \; u \leq \alpha(\tilde{\xi}_{1:T}, \xi^{(r-1)}_{1:T}) \\ \xi^{(r-1)}_{1:T} & \text{otherwise} \end{cases}$$

The simplest way to introduce negative correlation between the trajectories generated with the FFBS algorithm is to use, at a given iteration r of the sampler and for the t-th hidden state, a set of K uniform random numbers $U_{t,k}^{(r)}$, $k = 1, \ldots, K$ generated following the permuted displacement method (see Arvidsen and Johnsson [1982] and Craiu and Meng [2005]) given in Algorithm 4. The uniform random numbers are then use within the BS procedure to generate correlated proposals.

Algorithm 4 Permuted displacement method

- Draw $r_1 \sim \mathcal{U}_{[0,1]}$
- For $k = 2, \ldots, K-1$, set $r_k = \lfloor 2^{k-2} r_1 + 1/2 \rfloor$ where $\lfloor x \rfloor$ denotes the fractional part of x
- Set $r_K = 1 - \{2^{K-2} r_1\}$
- Pick at random $\sigma \in S_K$, where S_K is the set of all possible permutation of the integers $\{1, \ldots, K\}$
- For $k = 1, \ldots, K$, set $U_k = r_{\sigma(k)}$

For $K = 3$, Craiu and Meng [2005] show that the random numbers generated with the permuted displacement method are pairwise negatively associated (PNA). The definition of PNA given in the following is adopted from Craiu and Meng [2005].

Definition 1 (pairwise negative association)**.** The random variables $\xi_{t,1}$,$\xi_{t,2}$. . . ,$\xi_{t,K}$ are said to be pairwise negatively associated (PNA) if, for any nondecreasing functions

f_1, f_2 and $(i,j) \in \{1,\ldots,K\}^2$ such that $i \neq j$

$$Cov(f_1(\xi_{t,i}), f_2(\xi_{t,j})) \leq 0$$

whenever this covariance is well defined.

The proof for the case $K \geq 4$ is still an open issue. For this reason we consider in our algorithm $K \leq 3$. The presence of PNA in the case of $K \geq 4$ proposals depends on the degrees of uniformity of the filtering probability and the gain of efficiency should be proved computationally in each applications.

We use the permuted sampler to generate $K = 2$ multi-move and correlated proposals in the backward sampling step of the FFBS. In order to show how the antithetic sampler works, we consider the case where the hidden Markov switching process has two states, i.e. $\xi_t = (\xi_{1t}, \xi_{2t})'$ and for notational convenience let $\{q_t^{(r)}\}_{t=1,\ldots,T}$ be the sequence of filtered probabilities of being in state 1 at the r-th iteration of the sampler.

Proposition 3.3.1. *Define backward antithetic samples $\xi_{t,1}$ and $\xi_{t,2}$ as follows*

$$\xi_{t,1} = \begin{pmatrix} \mathbb{I}_{U_t^{(r)} < q_t^{(r)}} \\ \mathbb{I}_{U_t^{(r)} \geq q_t^{(r)}} \end{pmatrix}, \qquad \xi_{t,2} = \begin{pmatrix} \mathbb{I}_{V_t^{(r)} < q_t^{(r)}} \\ \mathbb{I}_{V_t^{(r)} \geq q_t^{(r)}} \end{pmatrix}$$

where $V_t^{(r)} = 1 - U_t^{(r)}$ and $U_t^{(r)} \sim \mathcal{U}_{[0,1]}$. Then it is possible to show by following the rules of expectation that

$$Cov(\xi_{t,1}^{(r)}, \xi_{t,2}^{(r)}) = \begin{pmatrix} (2q_t^{(r)} - 1)\mathbb{I}_{q_t^{(r)} > \frac{1}{2}} - \left(q_t^{(r)}\right)^2 & \left(q_t^{(r)}(1 - q_t^{(r)})\right)^2 \\ \left(q_t^{(r)}(1 - q_t^{(r)})\right)^2 & (1 - 2q_t^{(r)})\mathbb{I}_{q_t^{(r)} < \frac{1}{2}} - \left(1 - q_t^{(r)}\right)^2 \end{pmatrix}.$$

Proof The result follows from the application of expectation properties.

Using the expected value of the square of the Euclidean distance, $d(\xi_{t,1}, \xi_{t,2})$, between these two antithetic samples

$$E[d^2(\xi_{t,1}, \xi_{t,2})] = 2 - 2\left((2q_t^{(r)} - 1)\mathbb{I}_{q_t^{(r)} > \frac{1}{2}} + (1 - 2q_t^{(r)})\mathbb{I}_{q_t^{(r)} < \frac{1}{2}}\right), \tag{3.16}$$

it is possible to verify that extremely antithetic proposals are obtained when the distance on average is optimal. From equation (3.16) extreme antithetic sample obtained when $q_t^{(r)}$ is equal to 0.5, which can be easily found in applications where regimes exhibit similar persistence level.

3.3.2 Auxiliary models for defining the proposal distribution

In order to build up proposal distributions for the state variables, we will exploit all the knowledge we have about the full conditional distribution. The first step is to approximate the MS-GARCH model by eliminating the problem of path dependence and then derive a proposal distribution for state variables from the auxiliary model thus obtained. A possible way of circumventing the path dependence problem inherent in the MS-GARCH model is to replace the lagged conditional variance appearing in the definition of the GARCH model with a proxy. In the literature there are different auxiliary models which differ only by the content of the information used in the definition of the proxy introduced in each case. In general, various MS-GARCH (as available in the literature) can be obtained by approximating the conditional variance

$$\sigma_t^2(y_{1:t-1}, \theta_\sigma(s_t)) = V[y_t | y_{1:t-1}, s_{1:t}] = V[\epsilon_t | y_{1:t-1}, s_{1:t}]$$

of the GARCH process as follow

$$\sigma_t{}^2(y_{1:t-1}, \xi_t' \theta_\sigma) \approx \xi_t' \gamma + (\xi_t' \alpha)\epsilon^2_{(X)t-1} + (\xi_t' \beta)\sigma^2_{(X)t-1}. \tag{3.17}$$

In the subsection 3.3.2 - 3.3.2, we present alternative specifications of $\epsilon_{(X)t-1}$ and $\sigma^2_{(X)t-1}$ that define different approximations of the MS-GARCH model. The variable X can take on any of B, G, D, SK, K with each label representing basic, Gray [1996], Dueker [1997], simplified Klaassen [2002] and Klaassen [2002] approximations, respectively.

Basic model approximation (model B)

As a first attempt at eliminating the path dependence problem, we note that the conditional density of ϵ_t is a mixture of normal distributions with zero means and time varying variances. Hence, we approximate the switching GARCH model by replacing the lagged conditional variance, σ^2_{t-1}, with the variance $\sigma^2_{(B)t-1}$ of the conditional density of ϵ_t i.e.

$$\begin{aligned}
\epsilon_{(B)t-1} &= y_{t-1} - \mu_{(B)t-1} \\
\mu_{(B)t-1} &= E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-2}] = E[y_{t-1}|y_{1:t-2}] \\
&= \sum_{m=1}^{M} \mu_{t-1}(y_{1:t-2}, e'_m\theta_\mu) q(\xi_{t-1} = e_m|y_{1:t-2}), \\
\sigma^2_{(B)t-1} &= E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\sigma)|y_{1:t-2}] = E[\epsilon^2_{t-1}|y_{1:t-2}] = V(\epsilon_{t-1}|y_{1:t-2}) \\
&= \sum_{m=1}^{M} \sigma^2_{t-1}(y_{1:t-2}, e'_m\theta_\sigma) q(\xi_{t-1} = e_m|y_{1:t-2}).
\end{aligned}$$

Observe that in this approximation scheme $\mu_{(B)t-1}$ and $\sigma^2_{(B)t-1}$ are functions of $y_{1:t-2}$ and the information coming from y_{t-1} is lost. With $q(\xi_{t-1} = e_m|y_{1:t-2})$ known for $m = 1, \ldots, M$, $\mu_{(B)t-1}$ can easily be computed while $\sigma^2_{(B)t-1}$ can be computed recursively since $\sigma^2_{t-1}(y_{1:t-2}, e'_m\theta_\sigma)$ depends on $\sigma^2_{(B)t-2}$. Note that in this approximation the conditioning is on $y_{1:t-2}$. This approach represents a starting point for other approximations hence we tag it basic model approximation.

Gray's approximation (model G)

Gray [1996] notes that the conditional density of the observation process, y_t, of the switching GARCH model is a mixture of normal distributions with time-varying parameters. Hence, he suggests the use of the variance of the conditional density $\sigma^2_{(G)t-1}$ of y_t as a proxy for the lagged of the conditional variance σ^2_{t-1} switching GARCH process i.e.

$$
\begin{aligned}
\epsilon_{(G)t-1} &= y_{t-1} - \mu_{(G)t-1} \\
\mu_{(G)t-1} &= \mu_{(B)t-1} \\
\sigma^2_{(G)t-1} &= V(y_{t-1}|y_{1:t-2}) = V\left(E[y_{t-1}|y_{1:t-2}, \xi'_{t-1}]|y_{1:t-2}\right) + E[V\left(y_{t-1}|y_{1:t-2}, \xi'_{t-1}\right)|y_{1:t-2}] \\
&= V(\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-2}) + E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\sigma)|y_{1:t-2}] \\
&= E[(\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu))^2|y_{1:t-2}] - (E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-2}])^2 + \sigma^2_{(B)t-1} \\
&= \sum_{m=1}^{M}(\mu_{t-1}(y_{1:t-2}, e'_m\theta_\mu))^2 q(\xi_{t-1} = e_m|y_{1:t-2}) - (\mu_{(B)t-1})^2 + \sigma^2_{(B)t-1}.
\end{aligned}
$$

Similarly, as in the basic approximation, information on y_{t-1} is lost in this approximation scheme as $\mu_{(G)t-1}$ and $\sigma^2_{(G)t-1}$ are functions of $y_{1:t-2}$. By recursion, $\sigma^2_{(G)t-1}$ can be computed since $\sigma^2_{(B)t-1}$ depends on $\sigma^2_{(G)t-2}$ through $\sigma^2_{t-1}(y_{1:t-2}, e'_m\theta_\sigma)$. Within this framework the conditioning is also on $y_{1:t-2}$. The major difference between the basic model approximation and Gray's approximation can be seen from the development of the proxy i.e $V(\epsilon_{t-1}|y_{1:t-2})$ is replaced with $V(y_{t-1}|y_{1:t-2})$ in Gray's approximation.

Dueker's approximation (model D)

In the previous approximation schemes, the information coming from y_{t-1} is not used. Dueker [1997] suggests that y_{t-1} should be included in the conditioning set of the proxy while assuming that μ_{t-1} and σ^2_{t-1} are functions of $(y_{1:t-2}, \xi_{t-2})$. The following relation can thus be credited to him

$$\begin{aligned}
\epsilon_{(D)t-1} &= y_{t-1} - \mu_{(D)t-1} \\
\mu_{(D)t-1} &= E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-2}\theta_\mu)|y_{1:t-1}] = \sum_{m=1}^{M} \mu_{t-1}(y_{1:t-2}, e'_m\theta_\mu) q(\xi_{t-2} = e_m|y_{1:t-1}) \\
\sigma^2_{(D)t-1} &= E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-2}\theta_\sigma)|y_{1:t-1}] = \sum_{m=1}^{M} \sigma^2_{t-1}(y_{1:t-2}, e'_m\theta_\sigma) q(\xi_{t-2} = e_m|y_{1:t-1}).
\end{aligned}$$

The probability $q(\xi_{t-1} = e_m|y_{1:t})$ is a one period ahead smoothed probability which can be computed as:

$$\begin{aligned}
q(\xi_{t-1} = e_m|y_{1:t}) &= \sum_{i=1}^{M} q(\xi_{t-1} = e_m, \xi_t = e_i|y_{1:t}) \\
&= \sum_{i=1}^{M} q(\xi_{t-1} = e_m|\xi_t = e_i, y_{1:t}) q(\xi_i = e_i|y_{1:t}) \\
&= \sum_{i=1}^{M} q(\xi_{t-1} = e_m|\xi_t = e_i, y_{1:t-1}) q(\xi_i = e_i|y_{1:t}) \\
&= \sum_{i=1}^{M} \frac{q(\xi_{t-1} = e_m, \xi_t = e_i|y_{1:t-1}) q(\xi_i = e_i|y_{1:t})}{q(\xi_t = e_i|y_{1:t-1})} \\
&= q(\xi_{t-1} = e_m|y_{1:t-1}) \sum_{i=1}^{M} \frac{q(\xi_t = e_i|\xi_{t-1} = e_m, y_{1:t-1}) q(\xi_i = e_i|y_{1:t})}{q(\xi_t = e_i|y_{1:t-1})}
\end{aligned}$$

Within this framework we note that the conditioning is on $y_{1:t-1}$ while the functional form depends on $(y_{1:t-2}, \xi'_{t-2})$. We equally note that at every time step t the value of $q(\xi_{t-2} = e_m|y_{1:t-1})$ for all m is required for computation.

simplified Klaassen's approximation (model SK)

The following approximation is similar to Dueker's approximation (Model D). As opposed to Dueker's approximation, we assume that μ_{t-1} and σ^2_{t-1} are functions of $(y_{1:t-2}, \xi_{t-1})$. This modification leads to the following approximation Klaassen [2002] model.

$$
\begin{aligned}
\epsilon_{(SK)t-1} &= y_{t-1} - \mu_{(SK)t-1} \\
\mu_{(SK)t-1} &= E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-1}] = \sum_{m=1}^{M} \mu_{t-1}(y_{1:t-2}, e'_m\theta_\mu) q(\xi_{t-1} = e_m|y_{1:t-1}) \\
\sigma^2_{(SK)t-1} &= E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\sigma)|y_{1:t-1}] = \sum_{m=1}^{M} \sigma^2_{t-1}(y_{1:t-2}, e'_m\theta_\sigma) q(\xi_{t-1} = e_m|y_{1:t-1}).
\end{aligned}
$$

In the next approximation, the current regime will be added to the conditioning set of this version of the auxiliary model. Hence, this approximation will be identified as the simplified version of Klaassen [2002] model. In order to implement this approximation scheme the value of $q(\xi_{t-1} = e_m|y_{1:t-1})$ for all m is required at each point in time t.

Model 5: Klaassen's approximation (model K)

In each of the approximations described above, information relating to the current regime is ignored in the conditioning set. On observing this, Klaassen [2002] suggests the following approximation

$$
\begin{aligned}
\epsilon_{(K)t-1} &= y_{t-1} - \mu_{i,(K)t-1} \\
\mu_{i,(K)t-1} &= E[\mu_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\mu)|y_{1:t-1}, \xi_t = e_i] \\
&= \sum_{m=1}^{M} \mu_{t-1}(y_{1:t-2}, e'_m\theta_\mu) q(\xi_{t-1} = e_m|y_{1:t-1}, \xi_t = e_i) \\
\sigma^2_{i,(K)t-1} &= E[\sigma^2_{t-1}(y_{1:t-2}, \xi'_{t-1}\theta_\sigma)|y_{1:t-1}] \\
&= \sum_{m=1}^{M} \left(\mu^2_{t-1}(y_{1:t-2}, e'_m\theta_\mu) + \sigma^2_{t-1}(y_{1:t-2}, e'_m\theta_\sigma)\right) q(\xi'_{t-1} = e'_m|y_{1:t-1}, \xi'_t = e'_i) \\
&\quad - \left(\sum_{m=1}^{M} \mu_{t-1}(y_{1:t-2}, e'_m\theta_\mu) q(\xi_{t-1} = e_m|y_{1:t-1}, \xi_t = e_i)\right)^2,
\end{aligned}
$$

where

$$q(\xi_{t-1} = e_m | y_{1:t-1}, \xi_t = e_i) = \frac{q(\xi_{t-1} = e_m, \xi_t = e_i | y_{1:t-1})}{q(\xi_t = e_i | y_{1:t-1})}$$
$$= \frac{q(\xi_t = e_i | y_{1:t-1}, \xi_{t-1} = e_m) q(\xi_{t-1} = e_m | y_{1:t-1})}{q(\xi_t = e_i | y_{1:t-1})}.$$

Note that the computational complexity of this approximation is $\mathcal{O}(M^2T)$, as it requires the computation of $q(\xi_{t-1} = e_m | y_{1:t-1}, \xi_t = e_i)$ for all m and i at time step t.

3.3.3 Sampling θ

Sampling θ from the full conditional distribution will be done by separating the parameters of the transition matrix from the GARCH parameters, accordingly. We assume that the parameters of the transition probabilities are independent of GARCH parameters.

As regards the transition probability parameters, θ_π, their posterior distribution is given by the product of independent Dirichlet distributions

$$f(\theta_\pi | \xi_{1:T}, \theta_\mu, \theta_\sigma, y_{1:T}) = \prod_{m=1}^{M} \text{Dirichlet}(n_{1m} + \eta_{1m}, \ldots, n_{Mm} + \eta_{Mm}), \tag{3.18}$$

where

$$n_{ij} = \sum_{t=1}^{T} \xi_{jt-1} \xi_{it},$$

which can be simulate from directly.

Given a prior density $f(\theta_\mu, \theta_\sigma)$, the posterior density of the GARCH parameters, $(\theta_\mu, \theta_\sigma)$, can be expressed as

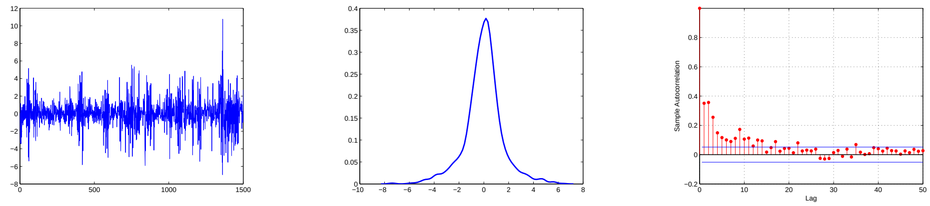
$$f(\theta_\mu, \theta_\sigma | \xi_{1:T}, \theta_\pi, y_{1:T}) \propto f(\theta_\mu, \theta_\sigma) \prod_{t=1}^{T} \mathcal{N}(\mu_t(y_{1:t-1}, \xi_t' \theta_\mu), \sigma_t^2(y_{1:t-1}, \xi_t' \theta_\sigma)). \tag{3.19}$$

For this step of the Gibbs sampler, we apply an adaptive Metropolis-Hastings (MH) algorithm, since the full conditional distribution is known to be non-standard. Details can be found, as required, in the Appendix 3.A.

3.4 Monte Carlo experiments on simulated data sets

We generate a time series of length 1500 from the data generating process (DGP) corresponding to the model defined by equations (3.3) and (3.4) for two regimes ($M = 2$), time invariant transition probabilities and switching conditional mean and variance. In the simulation exercise, we set $(\mu_1, \mu_2) = (0.06, -0.09)$, $(\gamma_1, \gamma_2) = (0.30, 2.00)$, $(\alpha_1, \alpha_2) = (0.35, 0.10)$, $(\beta_1, \beta_2) = (0.20, 0.60)$, and $\pi_{11} = 0.98$, $\pi_{22} = 0.96$. This parameter setting corresponds to the one in Bauwens et al. [2010] for a similar Monte Carlo exercise. A relatively higher and more persistent conditional variance as compared to the first regime GARCH equation is implied by the second regime GARCH equation. Also, the probability of staying in each regime is close to one. A typical series of length 1500 simulated from this DGP exhibits volatility clusters (see Fig. 3.1(a)). The kernel estimate of the unconditional density has heavy tails (see Fig. 3.1(b)) and the excess kurtosis is estimated to be 3.57. The autocorrelation function (ACF) of the square of the same series (Fig. 3.1(c)) is significant and this calls for the use of autoregressive volatility models.

For each hidden state sampling algorithm described in Section 3.3.1 and the auxiliary models presented in Section 3.3.2, we perform 10000 Gibbs iterations after convergence according to the Geweke's (Geweke [1992]) diagnostic by testing for the equality of means between the first 10% and the last 50% samples of the converging chain. It is well known that Gibbs sampler produces drawings with very high positive serial correlation for most parameters and results for 10000 samples may yield poor approximations. In view of this, we consider every 10th draw after convergence of



(a) Time series (b) Density approximation (c) Sample autocorrelation of squared returns

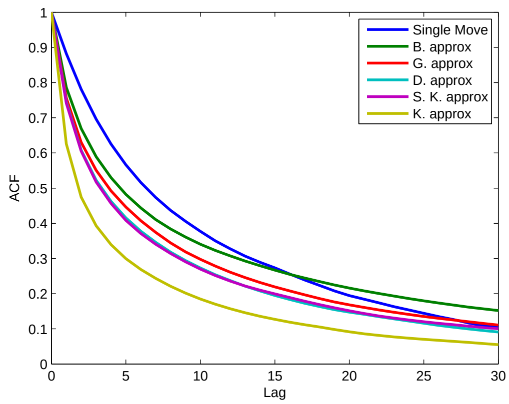
Figure 3.1: Simulated data for MS-GARCH model with parameter setting $(\mu_1, \mu_2) = (0.06, -0.09)$, $(\gamma_1, \gamma_2) = (0.30, 2.00)$, $(\alpha_1, \alpha_2) = (0.35, 0.10)$, $(\beta_1, \beta_2) = (0.20, 0.60)$, and $\pi_{11} = 0.98$, $\pi_{22} = 0.96$.

the Gibbs chain when evaluating the results presented in this section. This approach reduces the effective sample size to 1000 samples. The MCMC exercise is carried out by setting the initial state trajectory and parameter values of the algorithm to the maximum likelihood estimates of MS-GARCH model based on basic model approximation (see Section 3.3.1). The hyperparameters, ν_{ij} for $i, j = 1, 2$, of the prior distributions of the transition probabilities are set equal to 1. The support for other parameters are defined to obey their respective constraints. The case of two trials, $(K = 2)$, is considered within the different multi-point sampling strategies discussed earlier.

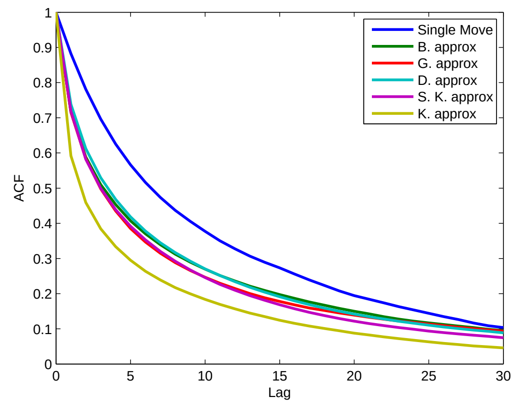
3.4.1 Comparison of algorithms

Using the first 10000 draws of the MCMC algorithm, we compare the efficiency of the different multi-move hidden state sampling algorithms against the single-move approach by computing the first 30 autocorrelations of the volatility, σ_t, for all time t. We summarize the results by taking the average of these autocorrelations over time. In Fig. 4.7 and 3.3 we plot the average of the autocorrelations for each of the multi-move multi-point sampling algorithms and single-move sampling scheme.

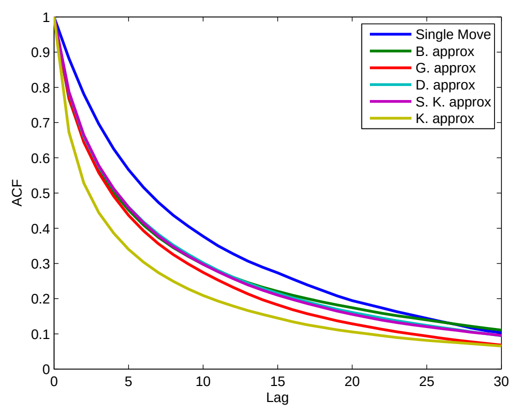
From Fig. 4.7, it may be deduced that under each multi-move multi-point sampling scheme Klaassen [2002] approximation produces the lowest autocorrelations



(a) Multiple-Try Metropolis (MTM)



(b) Multiple-trial Metropolized Independent Sampler (MTMIS)



(c) Multiple Correlated-Try Metropolis (MCTM)

Figure 3.2: Average autocorrelation of the volatility for each sampler (different plots) and approximation methods (different graph in each plot).
Note: B. approx, G. approx, D. approx, S. K. approx and K. approx respectively represent Basic model approximation (subsection 3.3.2), Gray's approximation (subsection 3.3.2), Dueker's approximation (subsection 3.3.2), simplified Klaassen's approximation (subsection 3.3.2) and Klaassen's approximation (subsection 3.3.2)

while, as expected, the single-move algorithm produces the highest autocorrelations. There is no clear difference among the autocorrelations produced by other approximations in each of the sampling scheme. On the other hand, from Fig. 3.3 we observe that MTMIS consistently outperform other sampling schemes. The efficiency of the various multi-move multi-point samplers are further assessed by computing the inefficiency factor and the relative inefficiency of the multi-point algorithms.

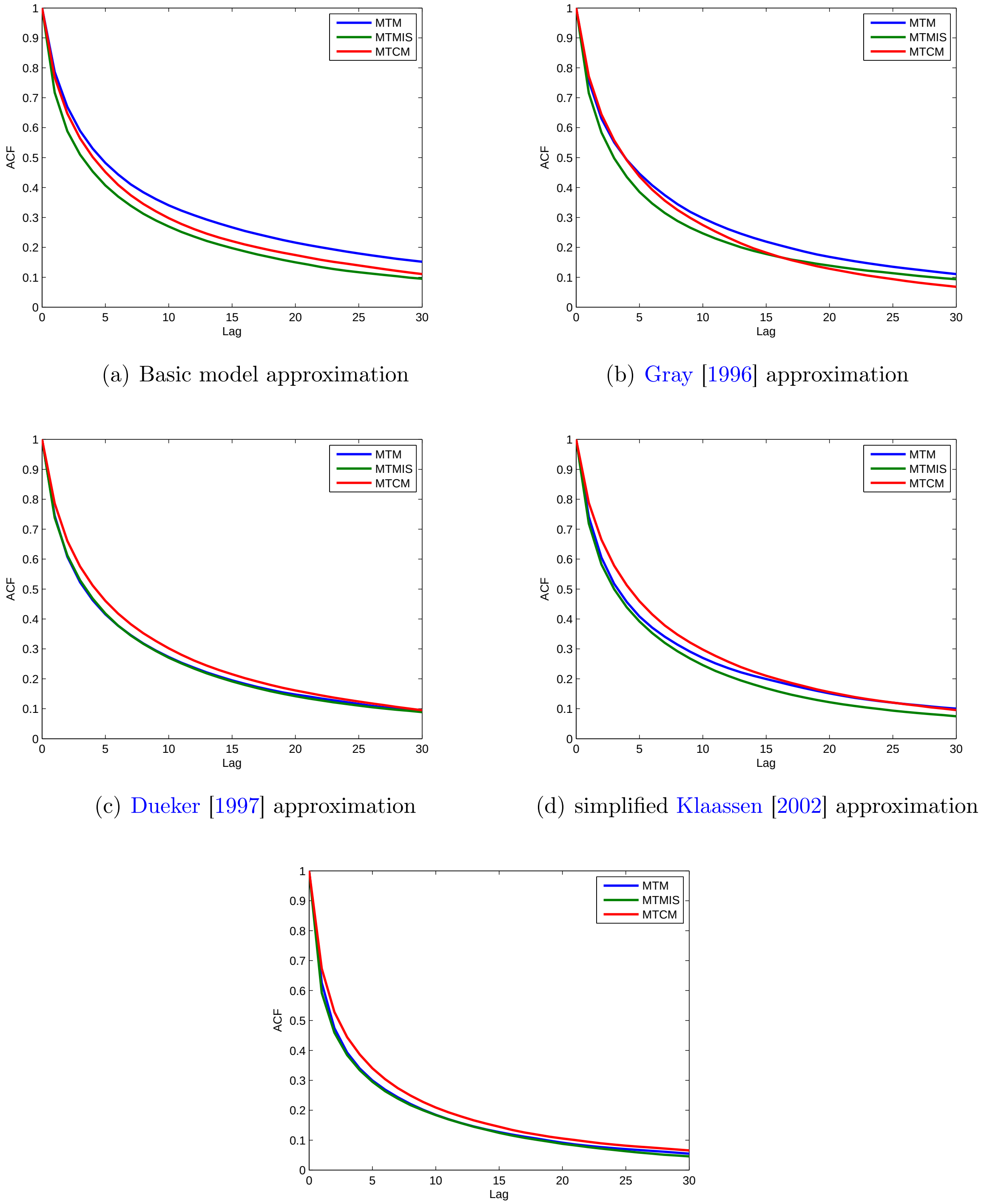
Let $\sigma_t^{(1)}, \ldots, \sigma_t^{(G)}$ denote a sample from the posterior distribution of a random variable σ_t. Then inefficiency factor (IF) is

$$IF = 1 + 2\sum_{l=1}^{L} w_l \rho_l, \tag{3.20}$$

where ρ_l, $l = 1, 2, \ldots, L$ is the autocorrelation function of $\sigma_t^{(1)}, \ldots, \sigma_t^{(G)}$ at lag l and w_l is the associated weight (see Robert and Casella [2007] for details). If the samples are independent then $IF = 1$. We define the relative inefficiency (RI) factor of two competing algorithms A and B with inefficiency factor IF_A and IF_B respectively, as

$$RI = \frac{Time_A}{Time_B} \times \frac{IF_A}{IF_B}, \tag{3.21}$$

where $Time_A$ and $Time_B$ are the computing times of each algorithm. RI measures the factor by which the run-time of algorithm A must be increased to achieve algorithm B's precision; values greater than one suggest that algorithm B is more efficient. In Tab. 3.1 we present the IF and report the RI for various multi-move multi-point algorithms relative to the single-move sampling technique. The table provides results for the two σ_t variables with the highest single-move inefficiency factor, i.e. σ_{513} and σ_{514} and the average σ over the volatilities. The number of lags over which we calculate both the IF and RI is set equal to $L = 500$. The inefficiency factor shows that MTMIS generally performs best among the sampling techniques while Klaassen [2002] is the best among the approximation methods (see Tab. 3.1). We



(a) Basic model approximation

(b) Gray [1996] approximation

(c) Dueker [1997] approximation

(d) simplified Klaassen [2002] approximation

(e) Klaassen [2002] approximation

Figure 3.3: Average autocorrelation of the volatility for each approximation method (different plots) and samplers (different line in each plot).

equally observe that our multi-move multi-point algorithms are more efficient than the single-move sampling technique for the state variable (see columns RI in Tab. 3.1). In order to achieve the precision of the multi-point sampling schemes, results from Tab. 3.1 suggest that, on average, one needs to run the single-move algorithm 72 times longer. Similarly, we observe from Tab. 3.1 that MTMIS under Klaassen's approximation is the most efficient combination. It does not come too much as a surprise that there is no appreciable difference in the efficiency of the MTCM over standard MTM because of the small number of multi-point used (see Craiu and Lemieux [2007] for discussion).

In Tab. 3.2, we report the posterior means and standard deviations of the parameters and the transition probabilities of the MS-GARCH using the various sampling schemes and approximation methods. With the exception of a few cases, the parameter posterior means obtained using our multi-move multi-point sampling schemes have more values within one posterior standard deviation away from the true values. We also quantify the performance of these estimates by calculating the mean squared error (MSE) of the posterior means of parameters relative to the true parameters i.e.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - \theta_i)^2, \tag{3.22}$$

where n is the number of parameters, $\hat{\theta}_i$ is the parameter estimate of the i-th element, θ_i, of the DGP parameter set.

The results in Tab. 3.2 show that some of the estimates are a bit off their respective true parameters. A possible explanation for this bias could be related to the variance scaling factor of the proposal distribution for sampling the GARCH parameters, which needs to be fine tuned in order to obtain estimates closer to the true parameters. Another explanation may be related to the choice of identification constraint. In our simulation exercise we considered a simple ordering of the inter-

Table 3.1: Inefficiency (IF) and relative inefficiency (RI) factors.

	Single-Move	MTM		MTMIS		MTCM	
	IF	IF	RI	IF	RI	IF	RI
B. approx.							
σ_{513}	7.62	5.39	53.38	5.03	59.00	5.33	62.02
σ_{514}	7.51	5.49	51.65	5.07	57.67	5.45	59.85
σ	5.68	5.08	42.20	4.68	47.27	5.05	48.90
G. approx.							
σ_{513}	7.62	5.37	53.38	5.29	56.05	5.05	65.73
σ_{514}	7.51	5.33	53.00	5.18	56.41	4.91	66.57
σ	5.68	5.02	42.55	4.66	47.47	4.92	50.23
D. approx.							
σ_{513}	7.62	4.88	58.73	4.94	59.15	5.33	61.24
σ_{514}	7.51	4.91	57.60	4.88	58.98	5.33	60.30
σ	5.68	4.73	45.17	4.73	45.97	5.11	47.56
S. K. approx.							
σ_{513}	7.62	4.87	59.85	4.51	65.89	5.68	58.56
σ_{514}	7.51	4.73	60.73	4.57	64.18	5.80	56.55
σ	5.68	4.77	45.49	4.43	50.01	5.00	49.55
K. approx.							
σ_{513}	7.62	3.09	90.18	3.07	92.15	3.97	79.53
σ_{514}	7.51	3.02	90.73	3.04	91.53	4.00	77.69
σ	5.68	3.77	55.07	3.70	56.96	4.08	57.66

Notes: IF and RI as defined in 3.20 and 3.21 respectively are computed using 10000 draws; the two σ_t variables shown are those for which the single move algorithm gives the largest IF; σ is the average over all σ_t; B. approx, G. approx, D. approx, S. K. approx and K. approx respectively represent Basic model approximation (3.3.2), Gray's approximation (3.3.2), Dueker's approximation (3.3.2), simplified Klaassen's approximation (3.3.2) and Klaassen's approximation (3.3.2)

cept ($\gamma_1 < \cdots < \gamma_M$) of the MS-GARCH model in order to identify the regimes. This approach, as illustrated in Marin et al. [2005] and Jasra et al. [2005], may have some consequences (such as large bias) on the resulting inference. In view of this, we leave for further research the effect of other choices of identification constraint such as the random permutation MCMC sampling of Frühwirth-Schnatter [2006] and the re-labelling algorithm of Jasra et al. [2005] on the parameter estimates. Perhaps, a different choice of the identification constraint may bring about reduction

Table 3.2: Estimated posterior mean and standard deviation of MS-GARCH parameters using various state proposal distributions and multi-move algorithms.

	DGP	Single-	Multi-move B. approx.			Multi-move G. approx.			Multi-move D. approx.			Multi-move S. K. approx.			Multi-move K. approx.		
	Values	Move	MTM	MTMIS	MCTM	MTM	MTMIS	MCTM	MTM	MTMIS	MCTM	MTM	MTMIS	MCTM	MTM	MTMIS	MCTM
π_{11}	0.980	0.999	0.987	0.987	0.987	0.986	0.987	0.987	0.986	0.987	0.987	0.987	0.987	0.987	0.987	0.986	0.986
		(0.001)	(0.004)	(0.005)	(0.004)	(0.005)	(0.004)	(0.004)	(0.005)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
π_{22}	0.960	0.996	0.961	0.960	0.961	0.961	0.960	0.962	0.961	0.961	0.961	0.961	0.961	0.962	0.961	0.960	0.960
		(0.003)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.011)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)	(0.012)
μ_1	0.060	0.030	0.045	0.045	0.044	0.045	0.044	0.044	0.045	0.044	0.045	0.044	0.044	0.044	0.045	0.045	0.045
		(0.019)	(0.020)	(0.020)	(0.020)	(0.020)	(0.019)	(0.019)	(0.019)	(0.020)	(0.020)	(0.019)	(0.020)	(0.020)	(0.020)	(0.020)	(0.020)
μ_2	−0.090	−0.014	−0.151	−0.155	−0.150	−0.151	−0.153	−0.152	−0.155	−0.156	−0.155	−0.152	−0.153	−0.154	−0.154	−0.156	−0.154
		(0.012)	(0.094)	(0.094)	(0.095)	(0.092)	(0.092)	(0.090)	(0.094)	(0.093)	(0.095)	(0.094)	(0.094)	(0.094)	(0.092)	(0.097)	(0.093)
γ_1	0.300	0.107	0.349	0.348	0.351	0.350	0.347	0.348	0.354	0.347	0.349	0.349	0.348	0.349	0.350	0.347	0.349
		(0.021)	(0.039)	(0.039)	(0.039)	(0.039)	(0.040)	(0.040)	(0.039)	(0.039)	(0.038)	(0.041)	(0.040)	(0.039)	(0.038)	(0.040)	(0.039)
γ_2	2.000	0.190	2.770	2.883	2.851	2.923	2.877	3.005	2.855	2.993	2.901	2.992	2.959	2.895	2.907	2.982	2.853
		(0.042)	(1.110)	(1.132)	(1.155)	(1.138)	(1.078)	(1.147)	(1.137)	(1.099)	(1.127)	(1.160)	(1.070)	(1.088)	(1.156)	(1.125)	(1.072)
α_1	0.350	0.327	0.411	0.405	0.410	0.402	0.401	0.415	0.406	0.404	0.404	0.402	0.405	0.404	0.404	0.405	0.400
		(0.036)	(0.066)	(0.069)	(0.068)	(0.067)	(0.067)	(0.067)	(0.070)	(0.067)	(0.064)	(0.064)	(0.064)	(0.065)	(0.065)	(0.067)	(0.066)
α_2	0.100	0.377	0.111	0.108	0.109	0.111	0.112	0.112	0.109	0.110	0.110	0.107	0.111	0.111	0.110	0.111	0.112
		(0.083)	(0.055)	(0.057)	(0.055)	(0.055)	(0.055)	(0.058)	(0.057)	(0.056)	(0.057)	(0.053)	(0.058)	(0.054)	(0.057)	(0.057)	(0.056)
β_1	0.200	0.657	0.099	0.103	0.098	0.103	0.106	0.100	0.094	0.105	0.098	0.103	0.106	0.102	0.101	0.105	0.104
		(0.034)	(0.057)	(0.060)	(0.059)	(0.059)	(0.062)	(0.059)	(0.055)	(0.058)	(0.059)	(0.061)	(0.059)	(0.059)	(0.059)	(0.061)	(0.059)
β_2	0.600	0.514	0.473	0.460	0.462	0.448	0.457	0.439	0.463	0.440	0.453	0.444	0.444	0.454	0.457	0.442	0.459
		(0.074)	(0.177)	(0.176)	(0.183)	(0.182)	(0.166)	(0.178)	(0.182)	(0.176)	(0.177)	(0.181)	(0.170)	(0.173)	(0.181)	(0.174)	(0.176)
MSE		0.362	0.063	0.082	0.076	0.089	0.081	0.106	0.077	0.103	0.086	0.103	0.096	0.084	0.086	0.101	0.077

Notes: Columns 3 - 18: Posterior means and standard deviations (in bracket) for the parameters of the MS-GARCH model. Results are based on a single sample of 1500 observations from the DGP defined by equation (3.3) and (3.4); MSE is computed using equation (3.22) in bias. The large standard deviation observed for some of the parameters in the second regime may be attributed to the fact that the regime is less frequently visited. Estimates of our multi-point multi-move sampling schemes are also observed to be very close to each other. These observation is not too surprising because the different approximations are based on the same fundamental while the different sampling strategies are meant to improve the efficiency of the MCMC algorithm. Furthermore, the assumption of low persistence ($\alpha_m + \beta_m << 1$ for $m = 1, 2$) in each regime of the MS-GARCH model may also have contributed to obtaining close parameter estimates. As with maximization algorithms, MCMC parameter estimates may change when they operate close to the boundaries of the constrained domain i.e. when MS-GARCH are strongly persistent in each regime. Nevertheless, unlike the single-move state sampling algorithm that is prone to getting stuck in some local optimum which in turn may lead to parameter estimates that are far off from the true value (see Tab. 3.2), our multi-point multi-move state sampling schemes is capable of avoiding such scenario. The superiority of our multi-point sampling schemes over the single-move sampler is further confirmed by their low MSE (see Tab. 3.2).

Fig. 3.7-3.9 in Appendix 3.B report the posterior densities of the parameters obtained via the different sampling strategies. The shapes of the posterior densities are unimodal, thus ruling out the presence of label switching. From these figures, we observe that the single-move (blue line) algorithm produces density estimates of the parameters that are highly peaked and are mostly not concentrated around the true parameter values. This confirms that the single-move state sampling algorithm produces samples that are trapped in some local modes of the posterior distribution. The multi-point multi-move algorithms on the other hand produce parameter posterior histograms concentrated around the true values. As noted for the parameter estimates, there is no clear difference among the density plot obtained using the multi-point and multi-move proposals.

It may be argued, as it is often the case, that the starting values for our MCMC algorithm may influence the parameter estimates and the results reported in Tab. 3.2 may have occurred by chance. To rule out this issue, the previous experiment is repeated 50 times, thus generating 50 samples of size 1500 of the same DGP and repeating the MCMC estimation for each sample using the same set of starting values. The mean and standard deviation (in parenthesis) for each set of 50 parameter estimates are reported in Tab. 3.3. A look at this table does not indicate much difference from the result presented in Tab. 3.2.

Based on these 50 replications we compute the average burnin period required to achieve convergence based on Geweke's convergence criterion. Tab. 3.4 shows that on average Klaassen [2002] approximation seems to converge the fastest among the different approximation methods while MTMIS offers the shortest burnin period among the multi-point sampling schemes. We also observe in Tab. 3.4 that on average the single-move scheme converges very fast. Unfortunately, as discussed above the seemingly fast convergence rate of single-move algorithm suggest that the algorithm has been trapped at a local optimum. This observation makes the single-move sampling scheme unattractive.

The performance of our multi-move multi-point algorithms relative to the single-move strategy is further examined by computing the percentage of correctly specified regimes. To do this, for each of the 50 replications, we calculate the average of the Gibbs output on the state variables and then assign mean states greater than one-half to regime 2 (and regime 1 otherwise). We find out that the single-move technique is able to classify on average 55% of the data correctly while the multi-move multipoint samplers classified between 93% and 96% of the data correctly. The acceptance rate of the multi-move multi-point proposals varies between 10% and 35% with the highest arising from MTMIS sampling schemes characterised with a proposal distribution constructed using Klaassen's approximation.

Table 3.3: Estimated mean and standard deviation of the posterior mean of MS-GARCH parameters based on 50 replications of the experiment described in Section 4

	DGP	Single-	Multi-move B. approx.			Multi-move G. approx.			Multi-move D. approx.			Multi-move S. K. approx.			Multi-move K. approx.		
	Values	Move	MTM	MTMIS	MCTM	MTM	MTMIS	MCTM	MTM	MTMIS	MCTM	MTM	MTMIS	MCTM	MTM	MTMIS	MCTM
π_{11}	0.980	0.999	0.982	0.980	0.981	0.981	0.980	0.980	0.981	0.981	0.980	0.980	0.982	0.979	0.980	0.980	0.979
		(0.0002)	(0.005)	(0.006)	(0.005)	(0.005)	(0.007)	(0.007)	(0.006)	(0.005)	(0.006)	(0.006)	(0.004)	(0.007)	(0.006)	(0.006)	(0.008)
π_{22}	0.960	0.995	0.961	0.961	0.960	0.960	0.959	0.961	0.962	0.962	0.961	0.961	0.962	0.960	0.960	0.960	0.960
		(0.002)	(0.007)	(0.008)	(0.008)	(0.007)	(0.009)	(0.007)	(0.006)	(0.007)	(0.007)	(0.007)	(0.007)	(0.008)	(0.007)	(0.007)	(0.008)
μ_1	0.060	0.057	0.053	0.054	0.054	0.053	0.054	0.058	0.051	0.054	0.055	0.051	0.054	0.057	0.056	0.055	0.055
		(0.023)	(0.014)	(0.014)	(0.013)	(0.015)	(0.013)	(0.015)	(0.011)	(0.014)	(0.013)	(0.011)	(0.014)	(0.013)	(0.013)	(0.013)	(0.013)
μ_2	−0.090	−0.028	−0.156	−0.159	−0.150	−0.154	−0.152	−0.167	−0.171	−0.150	−0.153	−0.172	−0.154	−0.154	−0.154	−0.151	−0.152
		(0.013)	(0.092)	(0.086)	(0.086)	(0.090)	(0.091)	(0.084)	(0.090)	(0.087)	(0.086)	(0.088)	(0.091)	(0.085)	(0.085)	(0.083)	(0.084)
γ_1	0.300	0.116	0.301	0.297	0.283	0.298	0.294	0.302	0.290	0.294	0.294	0.288	0.300	0.295	0.293	0.292	0.297
		(0.032)	(0.037)	(0.037)	(0.046)	(0.039)	(0.041)	(0.043)	(0.028)	(0.042)	(0.041)	(0.025)	(0.039)	(0.041)	(0.038)	(0.038)	(0.041)
γ_2	2.000	0.191	3.193	3.078	3.203	3.195	3.004	2.827	3.091	3.139	2.981	3.242	3.071	2.931	3.051	3.038	2.862
		(0.056)	(0.547)	(0.646)	(0.498)	(0.649)	(0.798)	(0.868)	(0.706)	(0.626)	(0.730)	(0.610)	(0.866)	(0.757)	(0.669)	(0.719)	(0.946)
α_1	0.350	0.325	0.392	0.380	0.388	0.388	0.391	0.369	0.384	0.389	0.382	0.381	0.383	0.373	0.383	0.383	0.373
		(0.054)	(0.054)	(0.056)	(0.049)	(0.050)	(0.050)	(0.066)	(0.054)	(0.049)	(0.055)	(0.061)	(0.057)	(0.068)	(0.056)	(0.053)	(0.070)
α_2	0.100	0.386	0.119	0.117	0.119	0.121	0.116	0.109	0.117	0.124	0.121	0.118	0.123	0.120	0.119	0.120	0.114
		(0.074)	(0.040)	(0.040)	(0.031)	(0.039)	(0.041)	(0.033)	(0.040)	(0.038)	(0.036)	(0.039)	(0.038)	(0.037)	(0.034)	(0.037)	(0.044)
β_1	0.200	0.650	0.198	0.201	0.225	0.202	0.209	0.190	0.207	0.212	0.205	0.211	0.200	0.204	0.207	0.207	0.200
		(0.060)	(0.042)	(0.041)	(0.057)	(0.046)	(0.050)	(0.061)	(0.037)	(0.052)	(0.057)	(0.033)	(0.041)	(0.052)	(0.047)	(0.050)	(0.059)
β_2	0.600	0.561	0.372	0.394	0.381	0.370	0.408	0.442	0.388	0.382	0.412	0.362	0.392	0.419	0.400	0.405	0.436
		(0.088)	(0.096)	(0.111)	(0.093)	(0.114)	(0.155)	(0.153)	(0.117)	(0.110)	(0.125)	(0.091)	(0.140)	(0.138)	(0.118)	(0.125)	(0.179)
MSE		0.360	0.148	0.121	0.150	0.149	0.105	0.072	0.124	0.135	0.100	0.161	0.120	0.090	0.115	0.112	0.078

Notes: Columns 3 - 18: Mean and standard deviations (in bracket) of posterior means for the parameters of the MS-GARCH model. Results based on 50 replications, each of which consists of 1500 observations from the DGP defined by equations (3.3) and (3.4); MSE is computed using the means of the posterior means and applying equation (3.22)

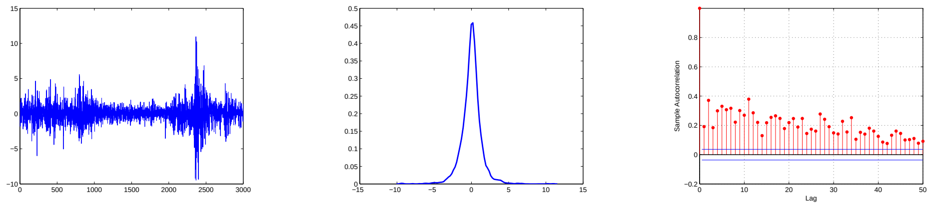
Table 3.4: Estimated burn-in period.

	Single-	Multi-move algorithms		
	Move	MTM	MTMIS	MCTM
models				
B. approx.	332	3550	2259	3648
G. approx.	332	1836	2090	3750
D. approx.	332	2579	2090	3667
S. K. approx.	332	5209	3113	5587
K. approx.	332	1208	2366	1816

Notes: Burn-in period is evaluated based on 10000 Gibbs iterations after convergence according to the Geweke's diagnostic. Estimated burn-in period is computed by taking the average burn-in period over 50 replications of the experiment.

Using our proposed multi-point sampling scheme we observe a substantial computational time reduction, as expected, when compared with the single-move scheme. We observe that the computational time required by our multi-point schemes are not different from one another. This observation is a consequence of the the two point sampling algorithm used in our MCMC algorithm. For one to observe an appreciable difference in the computational time among our multi-point sampling schemes, it will be required to increase the number of multi-point. Finally, comparing our MCMC algorithm in terms of their coding cost, the basic model approximation is the easiest approximation to code while the Klaassen's approximation is the most complex and challenging approximation to code. MTMIS on the other hand is the simplest multi-point sampling scheme among the set under consideration to code.

In summary, subject to the various performance criterion discussed above, Klaassen's approximation consistently performs at least as good as other MS-GARCH approximations used in constructing proposal distribution for sampling state trajectory. Be that as it may, the complexity of coding Klaassen's model limits its attractiveness. Subject to this and based on the fact that the parameter estimates obtained using the various multi-point multi-move algorithms are not significantly different from each other, we suggest the use of simplified Klaassen's



(a) Time series (b) Density approximation (c) sample autocorrelation of squared returns

Figure 3.4: Graphs for S&P 500 daily returns from 20/05/1999 to 25/04/2011.

approximation which is closely related to Klaassen's approximation and whose coding is not as complex as Klaassen's approximation for practical implementation. Among the multi-point sampling schemes we suggest that the MTMIS should be used for practical purposes since it is the most efficient and requires the least coding cost.

3.5 Empirical application: the S&P500 daily returns

We use our proposed Bayesian estimation procedure to fit MS-GARCH model to S&P500 daily percentage returns from 20/05/1999 to 25/04/2011[1] (3000 observations). The Bauwens et al. [2011] estimates would serve as a benchmark for evaluating the parameter estimates obtained using MTMIS algorithm coupled with the simplified Klaassen's approximation. Fig. 3.4 displays the returns sample path and kernel density estimate, and the autocorrelation of the squared returns. As one would expect from a typical financial time series, it exhibits strong persistence in the squared returns (see Fig. 3.4), slightly negative skewness and large excess kurtosis (see Tab. 3.5). These features calls for the use of a MS-GARCH model. We consider a two

Table 3.5: Descriptive statistics for S&P 500 daily returns.

Min.	max.	Mean	Std.	Skewness	Kurtosis
−9.470	10.960	−0.00022	1.353	−0.116	10.546.

[1]Thanks to Bauwens et al. [2011] for making the data available on https://sites.google.com/site/websiteofarnauddufays/

regime MS-GARCH model and set $\mu_1 = \mu_2 = 0$ because the sample mean of the series is close to zero. In Tab. 3.6, we report the posterior means and standard deviations (in bracket) from the estimation of the MS-GARCH model using MTMIS sampling scheme and simplified Klaassen's approximation for constructing the proposal distribution for sampling the state trajectory. These parameter estimates are based on 2000 samples i.e. we perform 20000 Gibbs iterations after convergence according to the Geweke's diagnostic and consider every 10th draw of this sample period to reduce the high correlation arising from the sampling algorithm. Fig. 3.5 displays the posterior density estimate for the distribution of the model parameters.

Table 3.6: Posterior means and standard deviations (S&P500 daily returns).

π_{11}	π_{22}	σ_1^2	σ_2^2	α_1	α_2	β_1	β_2
0.987	0.991	0.423	2.617	0.040	0.092	0.899	0.880
(0.003)	(0.002)	(0.006)	(0.015)	(0.011)	(0.013)	(0.018)	(0.016)

$\sigma_i^2 = \gamma_i/(1 - \alpha_i - \beta_i)$, $i = 1, 2$, represents the regime-specific unconditional variance

Our parameter estimates (see Tab. 3.6) are quite close to those obtained by Bauwens et al. [2011] with the same model specification on exactly the same set of data. We observe that our MS-GARCH model is characterized by low unconditional variance of 0.42 and persistence level of 0.93 for regime 1, while the second regime is characterized by high unconditional variance of 2.6 and persistence level of 0.97. In Fig. 3.6, we observe a sudden switch from low volatility to high volatility and then back to low volatility. The high volatility period corresponds to the 2008/2009 financial crisis. This observation shows that our model is flexible enough to capture extreme events confirming the empirical findings of Bauwens et al. [2011].

Finally, we assess the performance of our estimation procedure by computing the acceptance rate of the multi-point multi-move sampling scheme. We observed a very low acceptance rate of 0.1% on the empirical exercise as opposed to (10% − 35%) on the simulated data for the multipoint proposals. Despite the low acceptance rate for the multipoint proposals, we still have good results (as discussed above) considering our of single block sampling strategy for time series with many observations (1500 for simulated data and 3000 for empirical exercise) used. We expect that using the blocking scheme (as in So [2006]) the efficiency and the acceptance rate of our sampling procedure may increase. The issues of the block selction and block length are a subject of future research.

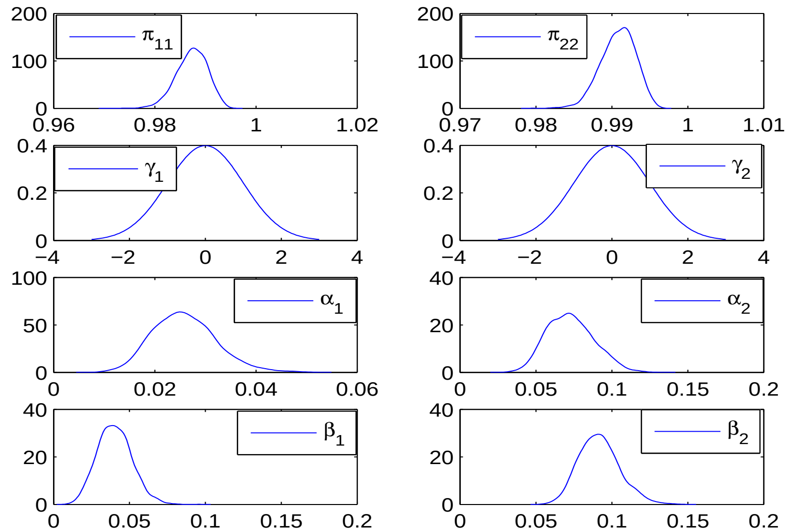


Figure 3.5: Posterior densities of the MS-GARCH parameters using multiple-trial Metropolized independent sampler (MTMIS) combined with the simplified Klaassen's approximation on S&P 500 daily returns.

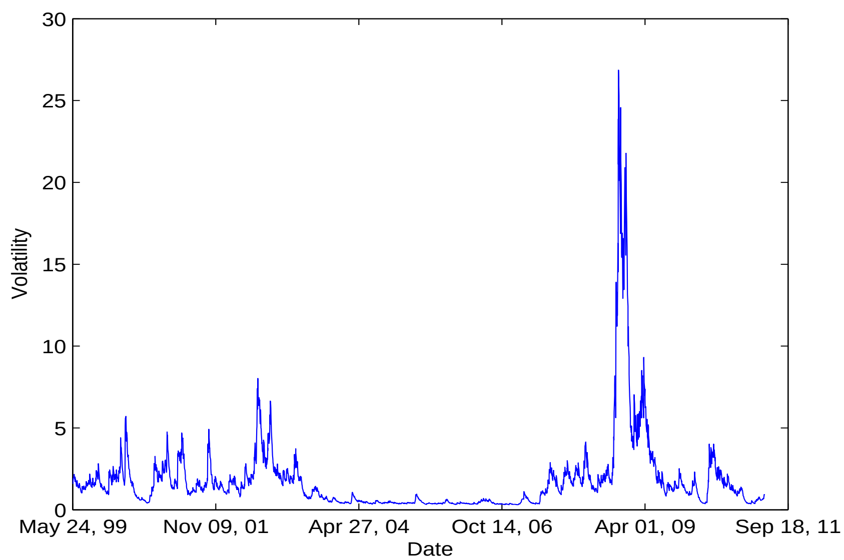


Figure 3.6: Conditional volatility estimated with the MS-GARCH model using multiple-trial Metropolized independent sampler (MTMIS) combined with the simplified Klaassen's approximation on S&P 500 daily returns.

3.6 Conclusion

In this paper, we deal with the challenging issue of designing efficient sampling algorithms for Bayesian inference on Markov-switching GARCH models. We provide some new generalized Metropolis algorithms based on the combination of multi-move and multi-points strategies. Our algorithms extend to Markov-switching nonlinear state space models the sampling algorithms propose by So [2006] for continuous nonlinear state space models.

More specifically, we apply the multiple-try sampling strategies of Craiu and Lemieux [2007] with a joint proposal distribution for the hidden states of the Markov-switching GARCH model. For generating candidate paths of the state variable we apply Forward Filtering Backward Sampling (FFBS) algorithm to an auxiliary MS-GARCH model. We propose different auxiliary models which are based on approximation of the GARCH conditional variance equation. We also design a multiple-try algorithm with correlated proposals. To this aim we introduce antithetic FFBS based on the permuted displacement method of Craiu and Meng [2005].

We compare our algorithms on simulated data and found that the multiple-trial Metropolized independent sampler in combination with Klaassen's approximation outperforms other multi-move multi-point sampling strategies under consideration. However, due to the cost of coding Klaassen's model and since there is no clear difference in the parameter estimates obtained by the various sampling algorithms, we suggest the use and apply the simplified Klaassen's approximation in combination with MTMIS for empirical purpose.

We conclude the paper with the indication of some future research lines. Our sampling approach can be extended by introducing a blocking scheme (as in So [2006]). We expect the efficiency and the acceptance rate of the resulting sampling procedure to be better. Also, the issue of the choice of block length could be a matter of future research. Another research line is related to the number of trial points required to achieve optimal performance of the multiple proposal scheme. Using insufficient trials may have negative effect on the performance of the multi-point algorithm. To this end, further research will be focused on computing the optimal choice of trial points. Finally, the effect of alternative prior distributions and identification constraints, on the parameter estimates, is a subject for further study.

‘

3.A Constructing proposal distribution for $\theta_\mu, \theta_\sigma$

Sample $\theta_\mu^{(r)}, \theta_\sigma^{(r)}$ from $f(\theta_\mu, \theta_\sigma | \xi_{1:T}^{(r)}, \pi^{(r)}, y_{1:T})$. Given a prior density $f(\theta_\mu, \theta_\sigma)$, the posterior density of $\theta_{\theta_\mu, \theta_\sigma}$ can be expressed as follows

$$f(\theta_\mu, \theta_\sigma | \xi_{1:T}^{(r)}, \pi, y_{1:T}) \propto f(\theta_\mu, \theta_\sigma) \prod_{t=1}^{T} \mathcal{N}(y_t; \xi_t^{(r)'} \mu, \sigma_t^2) \tag{3.23}$$

where,

$$\sigma_t^2 = \xi_t^{(r)'} \gamma + (\xi_t^{(r)'} \alpha)(y_{t-1} - \xi_{t-1}^{(r)\,'} \mu)^2 + (\xi_t^{(r)'} \beta) \sigma_{t-1}^2.$$

In order to generate $\theta_\mu, \theta_\sigma$ from the joint distribution we apply a further blocking of the Gibbs sampler i.e. We split the regime-dependent parameters in two subvectors, the parameter of the observation equation (θ_μ) and the parameters of the volatility process (θ_σ). For each subvector we implement a Metropolis-Hastings (MH) step that samples from a mixture of two normally distributed components in the case of θ_μ and mixture of two truncated normally distributed components in the case of θ_σ. The mixture is adapted during the burnin period. The expectation and the variance-covariance matrix of the first component are computed as described in Equation 3.24 and 4.38. This component behaves as an independent MH. We characterize the second component by a random walk proposal with the variance-covariance matrix specified as in the first component. We attach a low weight (0.05) to the first component and 0.95 to the second component.

As regards the parameters of the conditional expectation of the observation, we derive the mean and variance of the first component of the proposal distribution by considering an approximation of the full conditional distribution of μ,

$$f(\mu|\xi_{1:T}^{(r)}, \gamma^{(r-1)}, \beta^{(r-1)}, \alpha^{(r-1)}, y_{1:T}) \propto \prod_{t=1}^{T} \mathcal{N}(y_t; \xi_t^{(r)'}\mu, \sigma_t^2).$$

Given an approximation σ_t^{*2} of σ_t^2, it can easily be shown, by completing the square method, that the full conditional distribution of μ can be approximated by a normal distribution with mean and variance given by

$$m_\mu = \Sigma_\mu \left(\sum_{t=1}^{T} \frac{y_t \xi_t}{\sigma_t^{*2}} \right), \quad \Sigma_\mu = \left(\sum_{t=1}^{T} \frac{\xi_t \xi_t{}'}{\sigma_t^{*2}} \right)^{-1} \tag{3.24}$$

respectively, where

$$\sigma_t^{*2} = (\xi_t^{(r)'}\gamma^{(r-1)}) + (\xi_t^{(r)'}\alpha^{(r-1)})(y_{t-1} - \xi_{t-1}^{(r)}{}'\mu^{(r-1)})^2 + (\xi_t^{(r)'}\beta^{(r-1)})\sigma_{t-1}^{*2}.$$

The mean and variance thus constructed are used in defining the parameters of the mixture normal proposal distribution for μ.

As regards the parameters of the volatility process the full conditional is

$$f(\theta_\sigma|\xi_{1:T}^{(r)}, \mu^{(r)}, y_{1:T}) \propto \prod_{t=1}^{T} \mathcal{N}(y_t; \xi_t^{(r)'}\mu^{(r)}, \sigma_t^2). \tag{3.25}$$

We now follow the ARMA approximation of the MS-GARCH process i.e.

$$\begin{aligned}
\sigma_t^2 &= \xi_t'\gamma + (\xi_t'\alpha)\epsilon_{t-1}^2 + (\xi_t'\beta)\sigma_{t-1}^2 \\
\epsilon_t^2 &= \xi_t'\gamma + (\xi_t'\alpha + \xi_t'\beta)\epsilon_{t-1}^2 - (\xi_t'\beta)(\epsilon_{t-1}^2 - \sigma_{t-1}^2) + (\epsilon_t^2 - \sigma_t^2).
\end{aligned}$$

Let

$$w_t = \epsilon_t^2 - \sigma_t^2 = \left(\frac{\epsilon_t^2}{\sigma_t^2} - 1 \right) \sigma_t^2 = (\chi^2(1) - 1)\sigma_t^2$$

with

$$E_{t-1}[w_t] = 0; \quad \text{and} \quad Var_{t-1}[w_t] = 2\sigma_t^4.$$

Subject to the above and following Nakatsuma [1998] suggestion, we assume that $w_t \approx w_t^* \sim \mathcal{N}(0, 2\sigma_t^4)$. Then we have an "auxiliary"ARMA model for the squared error ϵ_t^2.

$$\begin{aligned} &\epsilon_t^2 = \xi_t'\gamma + (\xi_t'\alpha + \xi_t'\beta)\epsilon_{t-1}^2 - (\xi_t'\beta)w_{t-1}^* + w_t^*, \quad w_t^* \sim \mathcal{N}(0, 2\sigma_t^4) \\ \text{i.e.} \quad &w_t^* = \epsilon_t^2 - \xi_t'\gamma - (\xi_t'\alpha)\epsilon_{t-1}^2 - (\xi_t'\beta)(\epsilon_{t-1}^2 - w_{t-1}^*). \end{aligned} \tag{3.26}$$

Following Ardia [2008] we further express w_t^* as a linear function of $(3M \times 1)$ vector of $\theta_\sigma = (\gamma_1, \ldots, \gamma_M, \alpha_1, \ldots, \alpha_M, \beta_1, \ldots, \beta_M)'$. To do this, we approximate the function w_t^* by first order Taylor's expansion about $\theta_\sigma^{(r-1)} = (\gamma_1^{(r-1)}, \ldots, \gamma_M^{(r-1)}, \alpha_1^{(r-1)}, \ldots, \alpha_M^{(r-1)}, \beta_1^{(r-1)}, \ldots, \beta_M^{(r-1)})'$.

$$w_t^* \approx w_t^{**} = w_t^*(\theta_{-\pi}^{(r-1)}) - (\theta_\sigma - \theta_\sigma^{(r-1)})'\nabla_t \xi_t,$$

where

$$\nabla_t = -\begin{pmatrix} \frac{\partial w_t^*}{\partial \gamma_1} & 0 & \cdots & 0 & \frac{\partial w_t^*}{\partial \alpha_1} & 0 & \cdots & 0 & \frac{\partial w_t^*}{\partial \beta_1} & 0 & \cdots & 0 \\ 0 & \frac{\partial w_t^*}{\partial \gamma_2} & 0 & \vdots & 0 & \frac{\partial w_t^*}{\partial \alpha_2} & 0 & \vdots & 0 & \frac{\partial w_t^*}{\partial \beta_2} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\partial w_t^*}{\partial \gamma_M} & 0 & \cdots & 0 & \frac{\partial w_t^*}{\partial \alpha_M} & 0 & \cdots & 0 & \frac{\partial w_t^*}{\partial \beta_M} \end{pmatrix}'$$

and

$$\begin{aligned} \frac{\partial w_t^*}{\partial \gamma_k} &= -\xi_{tk} + (\xi_t'\beta)\frac{\partial w_{t-1}^*}{\partial \gamma_k} \\ \frac{\partial w_t^*}{\partial \alpha_k} &= -\xi_{tk}\epsilon_{t-1}^2 + (\xi_t'\beta)\frac{\partial w_{t-1}^*}{\partial \alpha_k} \\ \frac{\partial w_t^*}{\partial \beta_k} &= -\xi_{tk}(\epsilon_{t-1}^2 - w_{t-1}^*) + (\xi_t'\beta)\frac{\partial w_{t-1}^*}{\partial \beta_k} \end{aligned}$$

for $k = 1, \dots, M$, evaluated at $\theta_\sigma^{(r-1)}$.

Upon defining $r_t^* = w_t^*(\theta_{-\pi}^{(r-1)}) + \theta_\sigma^{\prime(r-1)} \nabla_t \xi_t$, it turns out that $w_t^{**} = r_t^* - \theta_\sigma' \nabla_t \xi_t$. Furthermore, by defining the $T \times 1$ vectors $\mathbf{w} = (w_1^{**}, \dots, w_T^{**})'$, $\mathbf{r}^* = (r_1^*, \dots, r_T^*)'$, a $T \times 3M$ matrix ∇ whose $t-$th row corresponds to $\xi_t' \nabla_t'$ as well as a $T \times T$ matrix

$$\mathbf{V} = 2 \begin{pmatrix} \sigma_1^{**4} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_T^{**4} \end{pmatrix},$$

with $\sigma_t^{**2} = (\xi_t^{(r)'} \gamma^{(r-1)}) + (\xi_t^{(r)'} \alpha^{(r-1)})(y_{t-1} - \xi_{t-1}^{(r)\,'} \mu^{(r)})^2 + (\xi_t^{(r)'} \beta^{(r-1)}) \sigma_{t-1}^{**2}$, we end up with $\mathbf{w} = \mathbf{r}^* - \nabla \theta_\sigma'$. Using this linear approximation, we can approximate the full conditional probability of the volatility parameters as

$$\begin{aligned} & f(\theta_\sigma | \xi_{1:T}^{(r)}, \mu^{(r)}, y_{1:T}) \propto \\ & \propto \frac{1}{|\mathbf{V}|^{\frac{1}{2}}} \exp\left(-\frac{\mathbf{w}'\mathbf{V}^{-1}\mathbf{w}}{2}\right) \mathbb{I}_{\{\gamma_1 > 0 \dots, \gamma_M > 0, 0 < \alpha_1 < 1, \dots, 0 < \alpha_M < 1, 0 < \beta_1 < 1 \dots, 0 < \beta_M < 1\}} \\ & \propto \mathcal{N}_{3M}(m_\sigma, \Sigma_\sigma) \mathbb{I}_{\{\gamma_1 > 0 \dots, \gamma_M > 0, 0 < \alpha_1 < 1, \dots, 0 < \alpha_M < 1, 0 < \beta_1 < 1 \dots, 0 < \beta_M < 1\}}, \end{aligned} \tag{3.27}$$

where

$$\begin{aligned} \Sigma_\sigma &= (\nabla' \mathbf{V}^{-1} \nabla)^{-1} \\ m_\sigma &= \Sigma \nabla' \mathbf{V}^{-1} \mathbf{r}^*. \end{aligned} \tag{3.28}$$

The mean and variance defined above are used to characterize proposal distribution for θ_σ, that is a mixture of truncated normal distributions. In our MCMC exercise, we sample from the normal mixture and check that each sample satisfies the constraints. Samples from the truncated distribution can be generated more efficiently by applying Gibbs sampling (see Fernández et al. [2007] for further details).

3.B Parameter posterior distributions

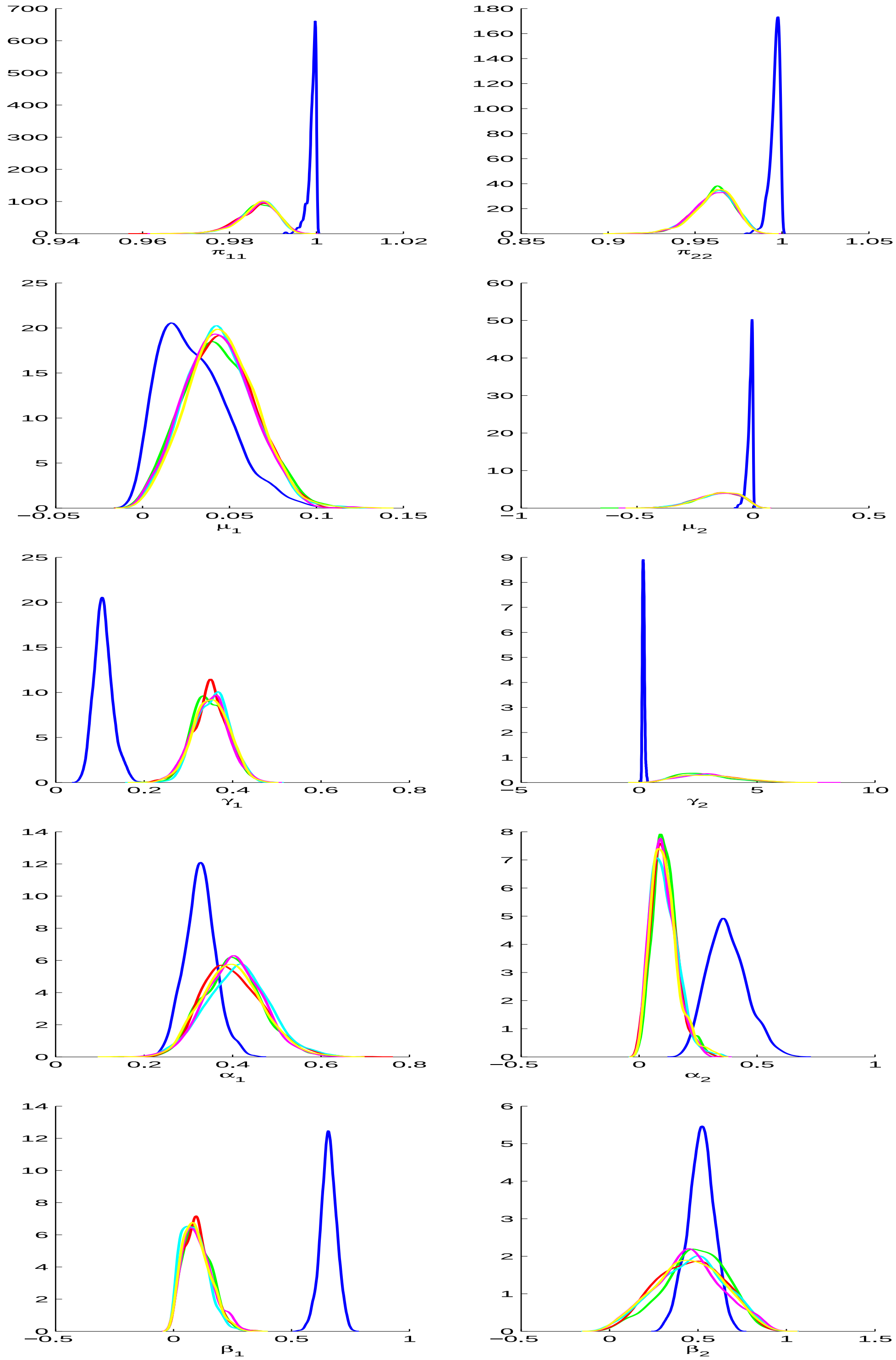


Figure 3.7: Posterior densities for the MS-GARCH model using the single-move algorithm and Multiple-Try Metropolis Sampler (MTM) algorithm.
Note: – single-move, – B. approx,– G. approx, – K. approx, – S. K. approx, – K. approx

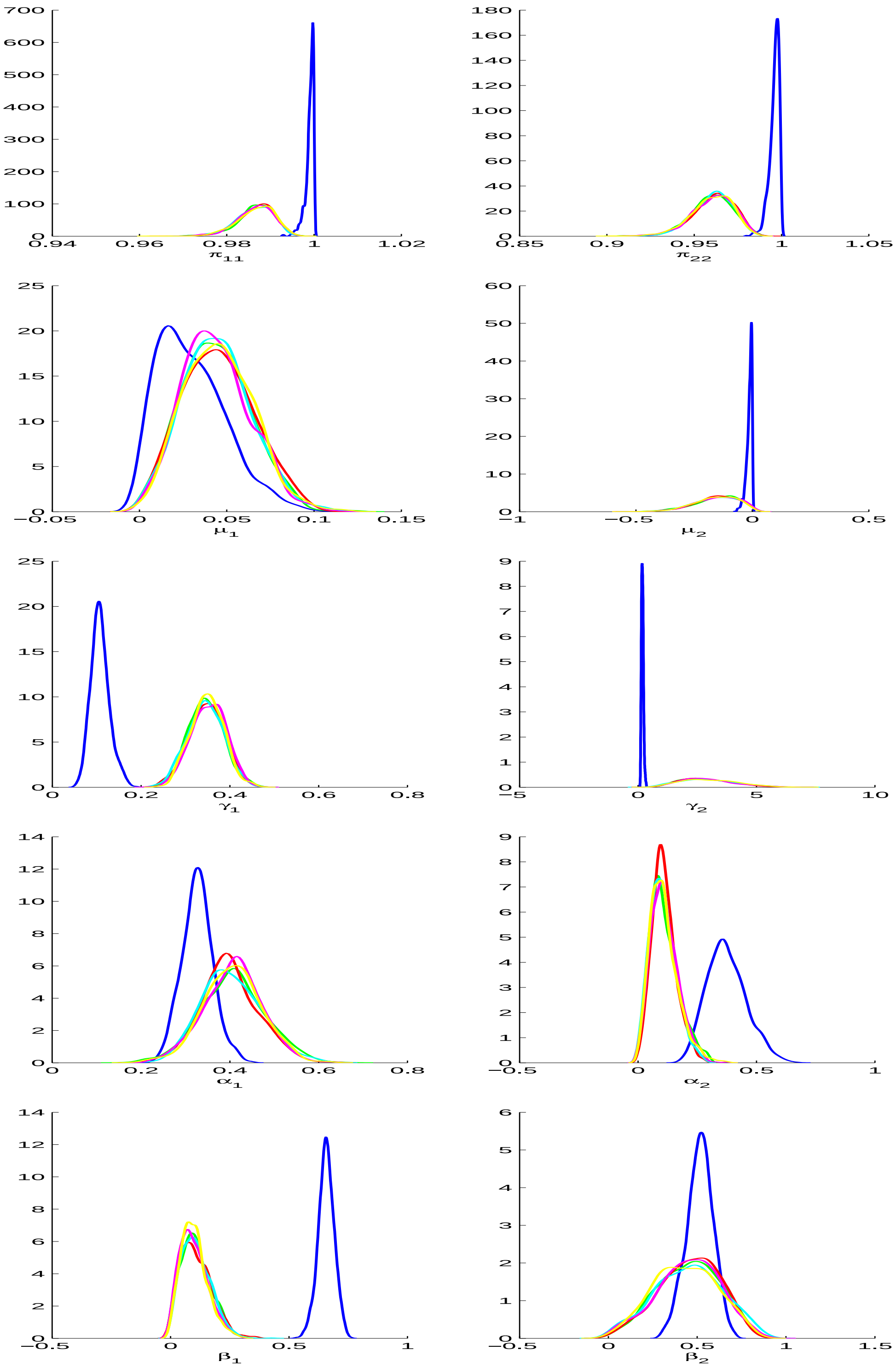


Figure 3.8: Posterior densities for the MS-GARCH model using the single-move and Multiple-trial Metropolized independent sampler (MTMIS) algorithms.
Note: – single-move, – B. approx,– G. approx, – K. approx, – S. K. approx, – K. approx

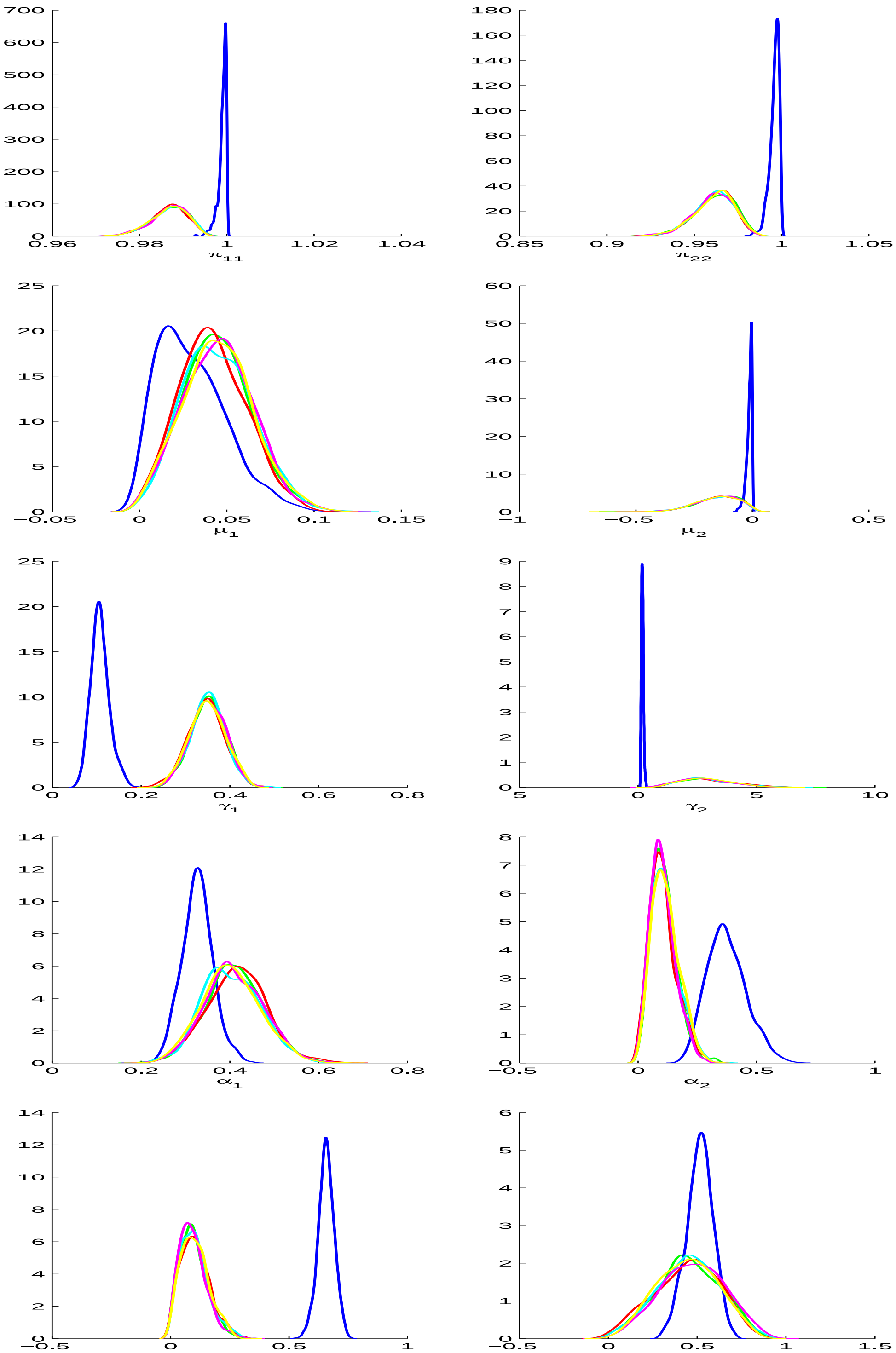


Figure 3.9: Posterior densities for the MS-GARCH model using the single-move and Multiple correlated-try Metropolis sampler (MTCM) algorithms.
Note: – single-move, – B. approx,– G. approx, – K. approx, – S. K. approx, – K. approx

Bibliography

A. Abramson and I. Cohen. On the stationarity of Markov-switching GARCH processes. *Econometric Theory*, 23:485–500, 2007.

D. Ardia. *Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications, volume 612 of Lecture Notes in Economics and Mathematical Systems.* Springer-Verlag, Berlin, Germany, 2008.

N. I. Arvidsen and T. Johnsson. Variance reduction through negative correlation - a simulation study. *J. of Statist. Comput. Simulation*, 15:119–127, 1982.

M. Augustyniak. Maximum likelihood estimation of the Markov-switching GARCH model. *Computational Statistics & Data Analysis*, 2013.

L. Bauwens, A. Preminger, and J. Rombouts. Theory and inference for a Markov switching GARCH model. *Econometrics Journal*, 13:218–244, 2010.

L. Bauwens, A. Dufays, and J. Rombouts. Marginal likelihood for Markov-switching and change-point GARCH. *CORE discussion paper, 2011/13*, 2011.

M. Billio, A. Monfort, and C.P. Robert. The simulated likelihood ratio method. *Doc. travail CREST, Insee, Paris*, 1998.

M. Billio, A. Monfort, and C. P. Robert. Bayesian estimation of switching ARMA models. *Journal of Econometrics*, 93:229–255, 1999.

S. Bizjajeva and J. Olsson. Antithetic sampling for sequential Monte Carlo methods with application to state space models. *Preprints in Mathematical Sciences, Lund University.*, 14:1 – 24, 2008.

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

J. Cai. A Markov model of switching-regime ARCH. *Journal of Business and Economics Statistics*, 12:309–316, 1994.

C. K. Carter and R. Kohn. On Gibbs sampling for state space models. *Biometrika*, 83:541–553, 1994.

R. Casarin, R. V. Craiu, and F. Leisen. Interacting multiple try algorithms with different proposal distributions. *Statistics and Computing*, 23(2):185–200, 2013.

S. Chib. Calculating posterior distributions and modal estimates in Markov mixture models. *Journal of Econometrics*, 75:79–97, 1996.

R. V. Craiu and C. Lemieux. Acceleration of the multiple-try Metropolis algorithm using antithetic and stratified sampling. *Statistics and Computing*, 17:109–120, 2007.

R. V. Craiu and X. L. Meng. Multi-process parallel antithetic coupling for forward and backward MCMC. *Ann. Statist.*, 33:661–697, 2005.

P. De Jong and N. Shephard. The simulation smoother for time series models. *Biometrika*, 82:339–350, 1995.

M. Dueker. Markov switching in GARCH processes in mean reverting stock market volatility. *Journal of Business and Economics Statistics*, 15:26–34, 1997.

A. Dufays. Infinite-state Markov-switching for dynamic volatility and correlation models. *CORE discussion paper, 2012/43*, 2012.

R. J. Elliott, J. W. Lau, H. Miao, and T. K. Siu. Viterbi-based estimation for Markov switching GARCH models. *Applied Mathematical Finance*, 19(3):1–13, 2012.

P. J. Fernández, P. A. Ferrari, and S. P. Grynberg. Perfectly random sampling of truncated multinormal distributions. *Advances in Applied Probabability 973-990.*, 39(4):973–990, 2007.

G. Fiorentini, C. Planas, and A. Rossi. Efficient MCMC sampling in dynamic mixture models. *Statistics and Computing*, 27:1–13, 2012.

S. Frühwirth-Schnatter. Data augmentation and dynamic linear models. *Journal of Time Series Analysis*, 15:183–202, 1994.

S. Frühwirth-Schnatter. *Mixture and Markov-switching Models*. Springer, New York, 2006.

J. Geweke. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In J. M. Bernardo, J. O. Berger, A. P. Dawid, and A. F. M. Smith, editors, *Bayesian Statistics 4*, pages 169–193. Oxford University Press, Oxford, 1992.

S. F. Gray. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42:27–62, 1996.

M. Haas, S. Mittnik, and M. Paolella. A new approach to Markov switching GARCH models. *Journal of Financial Econometrics*, 2:493–530, 2004.

J. D. Hamilton and R. Susmel. Autoregressive conditional heteroskedasticity and changes in regime. *Journal of Econometrics*, 64:307–333, 1994.

W. K. Hastings. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109, 1970.

Z. He and J.M. Maheu. Real time detection of structural breaks in GARCH models. *Computational Statistics and Data Analysis*, 54(11):2628–2640, 2010.

J. S. Henneke, S. T. Rachev, F. J. Fabozzi, and N. Metodi. MCMC-based estimation of Markov switching ARMA-GARCH models. *Applied Economics*, 43(3):259–271, 2011.

C. Holmes and A. Jasra. Antithetic methods for Gibbs samplers. *Journal of Computational and Graphical Statistics*, 18(2):401 – 414, 2009.

A. Jasra, C. Holmes, and D.A. Stephens. Markov chain Monte Carlo methods and the label switching problem in Bayesian mixture modeling. *Statistical Science*, pages 50–67, 2005.

S. Kaufman and S. Frühwirth-Schnatter. Bayesian analysis of switching ARCH models. *Journal of Time Series Analysis*, 23:425–458, 2002.

C.J. Kim and C.R. Nelson. *State-Space Models with Regime Switching: Classical and Gibbs-Sampling Approaches with Applications*. MIT Press, 1999.

F. Klaassen. Improving GARCH volatility forecasts with regime switching GARCH. *Empirical Economics*, 27:363–394, 2002.

S. J. Koopman and J. Durbin. Fast filtering and smoothing for multivariate state space models. *Journal of Time Series Analysis*, 21:281–296, 2000.

C. Lamoureux and W. Lastrapes. Persistence in variance, structural change, and the GARCH model. *Journal of Business and Economics Statistics*, 8:225–234, 1990.

J. Liu, F. Liang, and W. Wong. The multiple-try method and local optimization in Metropolis sampling. *Journal of the American Statistical Association*, 95:121–134, 2000.

J. S. Liu. *Monte Carlo Strategies in Scientific Computing*. Springer, 2002.

J. Marcucci. Forecasting stock market volatility with regime-switching GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, 9(4):1558–3708, 2005.

Jean-Michel Marin, Kerrie Mengersen, and Christian P Robert. Bayesian modelling and inference on mixtures of distributions. *Handbook of statistics*, 25:459–507, 2005.

N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller. Equations of state calculations by fast computing machines. *J. Chem. Ph.*, pages 1087–1092, 1953.

T. Mikosch and C. Starica. Nonstationarities in financial time series, the long-range dependence, and the IGARCH effects. *Review of Economics and Statistics*, 86: 378–390, 2004.

T. Nakatsuma. A Markov-chain sampling algorithm for GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, 3(2):107–117, 1998.

M. K. Pitt and N. Shephard. Antithetic variables for MCMC methods applied to non-Gaussian state space models. In *Proceedings of the Section on Bayesian Statistical Science.* Papers presented at the annual meeting of the American Statistical Association, Chicago, IL, USA, August 4–8, 1996 and the International Society for Bayesian Analysis 1996 North American Meeting, Chicago, IL, USA, August 2–3, 1996, 1996.

C. Robert and G. Casella. *Monte Carlo Statistical Methods.* Springer, 2007.

M.P.K. So. Bayesian analysis of nonlinear and non-Gaussian state space models via multiple-try sampling methods. *Statistics and Computing*, 16:125–141, 2006.

L. Wasserman. Asymptotic inference for mixture models using data-dependent priors. *Journal of the American Statistical Association*, 62:159–180, 2000.

Chapter 4

Markov Switching GARCH models for Bayesian Hedging on Energy Futures Markets

Abstract We propose Bayesian Markov Switching GARCH models for dynamic hedging in energy futures markets and contribute to the existing literature in several ways. First, we introduce a system of simultaneous equations consisting of return dynamics on the hedged portfolio and futures. More specifically, we assume that both the mean and variance of the hedged portfolio are governed by two unobserved discrete state processes, while the futures dynamics is driven by a univariate hidden state process. The noise in both processes are characterized by MS-GARCH. This formulation has at least two practical and conceptual advantages:(1) easy interpretation of the model parameters e.g. regime specific hedge ratios are easily generated and identified; (2) it provides an avenue to analyze the contribution of the volatility dynamics and switching probabilities to the optimal hedge ratio. Another contribution is the application of expected utility framework combined with regime-switching models to define a robust minimum variance hedging strategy which accounts for model parameter uncertainty. In this sense, we extend to regime switching environment the work of Lence and Hayes [1994a] on. Thirdly, we extend the FFBS sampling techniques in chapter 3 for a univariate chain MS-GARCH(1,1) to a multi-chain multivariate MS-GARCH(1,1) environment. Finally, the hedging model is applied to crude oil spot and futures markets.

Keywords : Markov Switching, Hedge ratio, Energy futures, GARCH

4.1 Introduction

Hedging is an investment position taken by investors to mitigate the adverse effect arising from changes in the price of a companion investment. A crucial issue, which has been subject to both theoretical discussions and econometric specifications, is the determination of the optimal hedge ratio, i.e. the number of derivative contracts to buy (or sell) for each unit of the underlying asset on which the investor bears risk. See Chen et al. [2003] for a review. In this paper, we focus on the econometric model specification and estimation procedure of the optimal hedge ratio proposed by Johnson [1960] and called Minimum Variance (MV) hedge ratio.

The MV hedge ratio is defined as the ratio of the covariance between the underlying spot and futures returns to the variance of the futures return. To apply this optimum hedge ratio in practice, Ederington [1979] suggests regressing the underlying spot returns against the futures returns and to use the estimate of the slope as MV hedge ratio. This approach has been widely criticized on the grounds that some of the well known stylized facts about asset returns are ignored. For example, it is well known that asset returns are usually not strictly stationary. To this end and to improve hedging performance, time-varying hedge ratios are proposed in the literature.

Two main approaches have been developed in the literature to estimate time-varying MV hedge ratios. One approach involves the estimation of the conditional second order moment of the underlying and futures returns captured by Generalized Autoregressive Conditional Heteroscedasticity (GARCH) models. See Haigh and Holt [2002], Chang et al. [2010] among others for illustration. The later approach treats the hedge ratio as a time-varying regression coefficient and focuses on the estimation of such a parameter (Lee et al. [2006], Chang et al. [2010] e.t.c.). Note that this hedge ratio works by re-balancing the hedged portfolio on a period by period basis. This may involve huge transaction costs and therefore it may not be worthwhile using this particular instrument for hedging. Also, it has been well documented in the empirical literature that the class of GARCH models exhibit high persistence of conditional variance, i.e. the process is close to being nearly integrated. In view of this, a few authors allow the optimal hedge ratio to be state-dependent. In line with the first approach of estimating time-varying hedge ratios, Alizadeh et al. [2008], Lee and Yoder [2007a], Lee and Yoder [2007b] among others propose various regime-switching multivariate GARCH models. More precisely, due to the path dependence problem of MS-GARCH models, these authors implement the multivariate extension of Gray [1996] model with differing characterization of time-varying covariance matrix. While Gray's model is attractive, its analytical intractability is a drawback i.e. it cannot be derived using any standard analytical approximation technique. Still in the same framework, Sheu and Lee [2012] argues that the dependence of both the derivative and the spot on the same hidden state process might be inappropriate. Thus, the authors propose the use of multichain Markov regime switching GARCH (MCSG) model. Alizadeh and Nomikos [2004] on the other hand follow the second approach and estimate time-varying hedge ratio by specifying a Markov switching variance model.

In practice, the actual values of the constituent parameters in the optimal hedge ratio are unknown. In this respect, traditionally, given the underlying econometric model, optimal hedge ratio is estimated by replacing the unknown parameters by their corresponding estimates. This approach is referred to as the "plug-in" or Parameter Certainty Equivalent (PCE) principle in the literature. Generally, decision makers are left to provide, using any estimation technique, estimates of the model parameters and substitute directly for their respective actual parameter values in the theoretical model. One of the problems with this approach is that it completely ignores estimation risk. As discussed above, several alternative econometric specification has been suggested in the literature for estimating the optimal hedge ratio. It is however puzzling to observe large differences in the estimated MV hedge ratios using these various econometric models on the same commodity. This observation further suggest that it may be very costly if estimation risk is ignored. Another problem is that relevant non sample information (such as insider information or subjective prior) available to the hedger are discarded in the decision making process. Perhaps, a more reliable decision may be arrived at by accounting for non-sample information in our estimation exercise.

A convenient and most frequently used approach in the literature for handling the problem of estimation risk is the expected utility paradigm. It may be argued that a rational decision maker would choose an action that maximizes its expected utility over the unknown parameter space. Early studies on this problem is pursued in Raiffa and Schlaifer [1961] and DeGroot [2005], among others. A review of the application of this theory, prior to 1978, to portfolio choice is provided in Bawa et al. [1979]. More recent application of this theory can be found in Kan and Zhou [2007]. As appealing as the expected utility theory sounds, it is laden with a number of computational issues. In many empirical application, analytical solution to either the optimization exercise and/or the integration problem are often not achievable. In view of this, alternative solution approach such as approximation or simulation is called for. Müller et al. [2004], Müller et al. [2004], among others, propose simulation based approach to the expected utility optimization problem.

Our contribution to the literature on time-varying hedge ratio is manifold. First, we propose a robust hedging ratio that accounts not only for parameter uncertainty but also for different state of the market. We follow a Bayesian decision rule to account for parameter uncertainty in the definition of optimal hedging strategies. See for example Lence and Hayes [1994a] and Lence and Hayes [1994b] for illustration. The second contribution is the use of MS-GARCH model in our econometric modeling framework. Our modeling approach differs from existing regime switching multivariate GARCH models in the hedging literature that nest within them both Gray's univariate regime switching model and state independent multivariate GARCH models. The closest approach to our modeling framework is the work by Alizadeh and Nomikos [2004]. We defer from Alizadeh and Nomikos [2004] in two ways. The first difference between ours and theirs lies in the characterization of the time-varying variance process. While Alizadeh and Nomikos [2004] considered a time-varying variance defined by an exponential function of the lagged 4-week moving average of the difference between the logarithm of the underlying and the logarithm of the futures, we consider a MS-GARCH model. The second difference relates to the properties of the underlying hidden process governing the observable processes. Alizadeh and Nomikos [2004] either assumes that the conditional variance of futures returns are regime independent or that the hidden process characterizing the dynamics of the hedged portfolio is independent of that influencing the futures returns process. We account for this limitations in our econometric framework. Examples of the application of MS-GARCH models in risk management can be found in Ardia [2008]. The third contribution involves a numerical procedure based on Monte Carlo Markov Chains (MCMC) for estimating the hedging ratio. We adopt a Bayesian approach for inferential purposes. The estimation exercise is carried out by following the sampling technique in Billio et al. [2012] to efficiently sample the state variable trajectory. In the empirical applications, we also efficiently estimate time-varying MV hedge ratios for crude oil and gasoline spot and futures prices used in Chang et al. [2010] and compare the result to conventional OLS method proposed by Ederington [1979].

The structure of this paper is as follows. In the next section, we present the conventional MV hedge ratio as well as the Bayesian hedging strategy. In Section 3, we illustrate with real data how the proposed model is implemented. Section 4 concludes the paper and provides suggestions for possible extensions.

4.2 Bayesian optimal hedge ratio

Let $(Y, \mathcal{Y}, P_\theta)$ be a probability observation space, with P_θ a parametric family of probability distributions and θ a parameter in the measurable parameter space $(\Theta, \mathcal{F}^\Theta)$. Let $\mathbf{y}_t = (RS_t, RF_t)' \in Y \subset \mathbb{R}^2$, $t = 1, \dots, T$, be an observable process, where RS_t, RF_t respectively corresponds to returns on the underlying and returns on the derivative (e.g., option, futures, etc.) at time t. Let us define the information set available at time t, as the σ-algebra $\mathcal{F}_t = \sigma(\{\mathbf{y}_s\}_{s \le t})$ generated by $\mathbf{y}_t$, $t = 1, \dots, T$ and denote with $\mathbf{y}_{s:t} = (\mathbf{y}_s, \dots, \mathbf{y}_t)$ a collection of observable variables.

Generally, consequent on the basic paradigm of expected utility theory and following standard hedging literature on commodities (e.g., see Haigh and Holt [2002] and references therein), at time t the optimal hedge ratio, h_t, is evaluated by solving the following optimization problem,

$$\arg\max_{h \in H} E(U|\mathcal{F}^\Theta_{t-1}) = \arg\max_{h \in H} \int_Y U(r(h, \mathbf{y}_t)) p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta) d\mathbf{y}_t, \tag{4.1}$$

where, $E(\cdot|\mathcal{F})$ is the conditional expectation operator, conditioning on a σ-algebra $\mathcal{F}$ $\mathcal{F}^\Theta_t = \sigma(\mathcal{F}_t \vee \mathcal{F}^\Theta)$ the information set generated by the collection of past values of observable process and prior information available about the parameter values, $U(\cdot)$ is the utility function, $r(h, \mathbf{y})$ is a function of decision variable, h, and a vector of random variables $\mathbf{y}$, H is the feasible set of hedge ratios, $p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \theta)$ is the joint probability density function (pdf) corresponding to $\mathbf{y}_t$ conditional on past values $\mathbf{y}_{1:t-1}$ and the parameter θ. For example, the minimum variance (MV) hedge ratio proposed by Johnson [1960] fits into this setting by assuming that: (i) the utility function is quadratic, and (ii) the function $r(h, \mathbf{y})$ is the returns on the hedged portfolio $(RS_t - hRF_t)$. Subject to these assumptions, minimizing the variance of the hedged portfolio reduces (4.1) to

$$h_t = \frac{Cov(RS_t, RF_t|\mathcal{F}^\Theta_{t-1})}{Var(RF_t|\mathcal{F}^\Theta_{t-1})}. \tag{4.2}$$

An important assumption in (4.1) is that $p(\mathbf{y}_t|\mathbf{y}_{1:t-1}\theta)$ is known with certainty. Unfortunately, in practice we are faced with incomplete knowledge either about (i) the functional form of $p(\mathbf{y}|\theta)$ (model risk) or (ii) the parameter vector, θ, given that the functional form of $p(\mathbf{y}_t|\mathbf{y}_{1:t-1},\theta)$ is known with certainty, (parameter uncertainty). We shall limit our discussion in what follows to the case characterized by parameter uncertainty while model risk problem is a subject for further research.

If the hedger does not know the true values of the parameters in θ, the problem as represented in (4.1) cannot be solved because $E(U|\mathcal{F}^{\Theta}_{t-1})$ is a function of these unknown parameters and therefore is also unknown. The classical solution to this problem follows the "plug-in" principle i.e. a point estimate $\hat{\theta} \in \mathcal{F}_{t-1}$ is substituted for the unknown parameter vector θ. Upon appropriate substitution, (4.1) becomes

$$\arg\max_{h\in H} E(U|\mathcal{F}_{t-1}) = \arg\max_{h\in H} \int_Y U(r(h,\mathbf{y}_t))p(\mathbf{y}_t|\mathbf{y}_{1:t-1},\hat{\theta})d\mathbf{y}_t. \tag{4.3}$$

In the technique described above, we act as if the parameters are known thus suggesting the name parameter certainty equivalent (PCE) as this technique is sometimes referred to in the literature. The uncertainty about the parameters in (4.1) are completely ignored in this approach. This calls for care when applying this method. Based on this, we adopt the Bayes' decision criterion (see Lence and Hayes [1994a]) by integrating out the unknown parameters in the product of $E(U|\mathcal{F}^{\Theta}_{t-1})$ and the posterior distribution of θ i.e.

$$\begin{aligned}
&\arg\max_{h\in H} E(E(U|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1}) = \\
&= \arg\max_{h\in H} \int_\Theta \left(\int_Y U(r(h,\mathbf{y}_t))p(\mathbf{y}_t|\mathbf{y}_{1:t-1},\theta)d\mathbf{y}_t \right) p(\theta|\mathbf{y}_{1:t-1})d\theta \\
&= \arg\max_{h\in H} \int_Y U(r(h,\mathbf{y}))p(\mathbf{y}_t|\mathbf{y}_{1:t-1})d\mathbf{y}_t,
\end{aligned} \tag{4.4}$$

where $p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$ is the marginal posterior predictive distribution. Unlike (4.1), (4.4) does not involve any unknown parameter, but requires some information about the parameters. The information can come from past values of the observation process or from other prior information included in $\mathcal{F}^{\Theta}$ and in the prior distribution of the parameters. In this case, the MV hedge ratio is

$$h_t^{BAY} = \frac{E(Cov(RS_t, RF_t|\mathcal{F}_{t-1}^{\Theta})|\mathcal{F}_{t-1})}{E(Var(RF_t|\mathcal{F}_{t-1}^{\Theta})|\mathcal{F}_{t-1})}. \tag{4.5}$$

As highlighted in Bawa et al. [1979], applying Bayes' criterion (4.4) in place of the PCE approach has at least three benefits. First, Bayes' criterion is supported by the basic axioms postulated by von Neumann-Morgenstern, whereas the PCE has no such axiomatic foundation. Second, all relevant (sample or non-sample) information about θ are taking into consideration through the posterior distribution in Bayes' method. In contrast, sample informations contained in the point estimates $\hat{\theta}$ are only needed to implement the PCE. Lastly, optimal average risk decision is arrived at by using Bayes' criterion.

In many situations, obtaining analytical solution to the Bayesian optimal hedge ratio problem in (4.4) can be a daunting task. This is because, in some cases, neither the maximization nor the integration exercise can be solved analytically, thus demanding for alternative approaches such as approximation and/or simulation based methods (see Müller [1999]). For example, the integrand may be too complex to integrate or the number of parameters to integrate over might be too large to evaluate analytically. In such a scenario, it is possible to approximate the optimization problem in (4.4) by using draws from the posterior distribution of θ given $\mathcal{F}_{t-1}$, which is a natural output of the MCMC approximation of the θ posterior distribution (see Amzal et al. [2006] and Müller et al. [2004]). Several alternative theories regarding the functional form of utility function (see for example Lence and Hayes [1994b] and

Haigh and Holt [2002]) other than the negative squared error loss function used in the MV hedge ratio exit in the literature for deriving the optimal hedge ratio. In this article, we shall limit our attention to the MV hedge ratio as it is the most commonly used optimal hedge ratio.

4.2.1 Econometric model specification

A popular econometric model used for calculating the optimal hedge ratio is the linear model proposed by Ederington [1979]. In this model a linear relationship is assumed between the underlying spot and futures returns

$$RS_t = \mu + \nu RF_t + \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} (0, \sigma^2), \tag{4.6}$$

where μ, ν and σ are the regression parameters. The ordinary least square (OLS) estimate of the coefficient of RF_t, ν, is then the MV hedging ratio. The assumption of constant variance and covariance in (4.6) implies time-invariant hedge ratio and thus makes this approach easy to implement. However, as highlighted by Myers [1991], this method fails to properly account for all relevant conditioning information available to hedgers when making their decision. Also, this method fails to account for some of the well known stylized facts, such as conditional heteroscedasticity and volatility clustering, commonly observed in financial data. In view of this and to allow for changes in the market conditions to affect the hedge ratios, Equation (4.6) is extended to an M state Markov switching model with a time-varying volatility process also characterized by regime switching.

Let us define two measurable state space $(X, \mathcal{X})$ and $(Z, \mathcal{Z})$ and unobserved processes, $s_t \in (X, \mathcal{X})$, $z_t \in (Z, \mathcal{Z})$, $t = 1, \ldots, T$, which are state variables representing, respectively, the state of the hedged portfolio and futures market at time t. Let $\mathcal{F}_t^X$

and $\mathcal{F}_t^Z$ be the sigma algebras generated respectively by s_u, and z_u, $u \leq t$. Then,

$$\begin{aligned} RS_t &= \mu(s_t) + \nu(s_t, z_t) RF_t + \sigma_t \eta_t, & \eta_t &\overset{iid}{\sim} \mathcal{N}(0,1), \\ \sigma_t^2 &= \gamma(s_t) + \alpha(s_t)\epsilon_{t-1}^2 + \beta(s_t)\sigma_{t-1}^2, \\ RF_t &= a(z_t) + \tau_t \zeta_t, & \zeta_t &\overset{iid}{\sim} \mathcal{N}(0,1), \\ \tau_t^2 &= \kappa(z_t) + \omega(z_t)\xi_{t-1}^2 + \psi(z_t)\tau_{t-1}^2, \end{aligned} \tag{4.7}$$

where, $\epsilon_t = \sigma_t \eta_t$, $\xi_t = \tau_t \zeta_t$, $\mu(s_t, z_t)$, $\nu(s_t)$, $\gamma(s_t) > 0$, $\alpha(s_t) \geq 0$, $\beta(s_t) \geq 0$, $a(z_t)$, $\kappa(z_t) > 0$, $\omega(z_t) \geq 0$, $\psi(z_t) \geq 0$, and $(s_t, z_t) \in \{1, \ldots, M\}^2$, $t = 1, \ldots, T$, is assumed to follow a $M \times M$-state first order Markov chain with transition probabilities $\{\pi_{ij,kl}\}_{i,j,k,l=1,2,\ldots,M}$:

$$\pi_{ij,kl} = p(s_t = i, z_t = j | s_{t-1} = k, z_{t-1} = l), \quad \sum_{i=1}^{M} \sum_{j=1}^{M} \pi_{ij,kl} = 1 \quad \forall\ k, l = 1, 2, \ldots, M. \tag{4.8}$$

The parameter shift functions $\mu(s_t)$, $\nu(s_t, z_t)$, $a(z_t)$, $\gamma(s_t)$, $\alpha(s_t)$, $\beta(s_t)$, $\kappa(z_t)$, $\omega(z_t)$ and $\psi(z_t)$ describe the dependence of parameters on the realized regimes s_t, z_t e.g.

$$\mu(s_t) = \sum_{i,j=1}^{M} \mu_i \mathbb{I}_{s_t=i} \mathbb{I}_{z_t=j}, \quad \nu(s_t, z_t) = \sum_{i,j=1}^{M} \nu_{ij} \mathbb{I}_{s_t=i} \mathbb{I}_{z_t=j}, \quad \text{with } \mathbb{I}_{s_t=i} = \begin{cases} 1, & \text{if } s_t = i \\ 0, & \text{otherwise,} \end{cases}$$

Let $s_{s:t} = (s_s, \ldots, s_t)$, $z_{s:t} = (z_s, \ldots, z_t)$, $(s, z)_{s:t} = \{(s_r, z_r)\}_{r=s:t}$, $RS_{s:t} = (RS_s, \ldots, RS_t)$, $RF_{s:t} = (RF_s, \ldots, RF_t)$ whenever $s < t$, $\theta_\pi = (\{\pi_{ij,kl}\}_{i,j,k,l=1,\ldots,M})$, $\theta_u^{RS} = (\mu_1, \ldots, \mu_M, \nu_{11}, \ldots, \nu_{MM})$, $\theta_a^{RF} = (a_1, \ldots, a_M)$, $\theta_\sigma = (\gamma_1, \ldots, \gamma_M, \alpha_1, \ldots, \alpha_M, \beta_1, \ldots, \beta_M)$, $\theta_\tau = (\kappa_1, \ldots, \kappa_M, \omega_1, \ldots, \omega_M\ , \psi_1, \ldots, \psi_M)$ and $\theta = (\theta_\pi, \theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau)$.

We summarize the theoretical implication of this extension on the optimal hedge ratio in the following proposition.

Proposition 4.2.1. *Suppose θ is known and assume that the observations are generated by the process described in (4.7). Then the conditional minimum variance hedge* *ratio at time t, is the solution to*

$$h_t = \arg\min_{h \in H} Var(RS_t - hRF_t|\mathcal{F}_{t-1}^{\Theta}) \tag{4.9}$$

which is given by

$$h_t = \frac{Cov(\mu(s_t), a(z_t)|\mathcal{F}_{t-1}^{\Theta})}{Var(RF_t|\mathcal{F}_{t-1}^{\Theta})} + \sum_{i,j=1}^{M} \nu_{ij} w_{ij}, \tag{4.10}$$

where

$$w_{ij} = \frac{\left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}_{t-1}^{\Theta}])\pi_{ij,..} p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta)\right)}{\sum_{i,j=1}^{M} \left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}_{t-1}^{\Theta}])\pi_{ij,..} p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta)\right)},$$

$$E(a(z_t)|\mathcal{F}_{t-1}^{\Theta}) = \sum_{(s,z)_{1:t-1}} \sum_{i,j=1}^{M} a_j \pi_{ij,..} p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta),$$

$$Cov(\mu(s_t), a(z_t)|\mathcal{F}_{t-1}^{\Theta}) = \sum_{(s,z)_{1:t-1}} \sum_{i,j=1}^{M} \left(\mu_i a_j - \mu_i E[a(z_t)|\mathcal{F}_{t-1}^{\Theta}]\right) \pi_{ij,..} p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta),$$

$$V(RF_t|\mathcal{F}_{t-1}^{\Theta}) = \sum_{i,j=1}^{M} \left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}_{t-1}^{\Theta}])\pi_{ij,..} p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta) \right),$$

$$\pi_{ij,..} = p(s_t = i, z_t = j | s_{t-1}, z_{t-1}, \theta),$$

and $\tau_t^2(k) = \kappa_k + \omega_k \xi_{t-1}^2 + \phi_k \tau_{t-1}^2$ *for* $k = 1, \ldots, M$ *and* $t = 1, \ldots, T$.

Proof: See Appendix (4.A).

Proposition 4.2.1 states that the optimal hedge ratio at any point in time can be determined by two components. The first one is given by the conditional covariance between the intercepts ($a(z_t)$ and $\mu(s_t)$) scaled by the conditional variance of RF_t. If $\nu(s_t, z_t) = 0$ and the spot and derivative rates go on average in the same direction within the same regime then the hedge ratio increases. The second component is a weighted average of the hedge ratios conditioning on the different states ($\nu_{ij}, i, j =$

$1,\ldots,M$). The weights are driven by the volatility of the returns on the derivative. This suggests that the dynamics of the variance process on the derivative plays an important role in estimating the MV hedge ratio. This is more evident in Remark 4.2.2 below.

Remark 4.2.2. If $a(z_t)$ is constant, then the optimal hedge ratio in (4.10) reduces to

$$h_t = \sum_{i,j=1}^{M} \nu_{ij} \left(\frac{\left(\sum_{(s,z)_{1:t-1}} \tau_t^2(j) p(s_t = i, z_t = j | s_{t-1}, z_{t-1}, \theta) p((s,z)_{1:t-1} | \mathcal{F}_{t-1}, \theta)\right)}{\sum_{i,j=1}^{M} \left(\sum_{(s,z)_{1:t-1}} \tau_t^2(j) p(s_t = i, z_t = j | s_{t-1}, z_{t-1}, \theta) p((s,z)_{1:t-1} | \mathcal{F}_{t-1}, \theta)\right)} \right). \tag{4.11}$$

Remark 4.2.3. If the dynamics of both the hedged portfolio and the derivative are govern by the same unobserved state process, s_t, then the optimal hedge ratio at time t is given by

$$h_t = \frac{Cov(\mu(s_t), a(s_t) | \mathcal{F}_{t-1}^{\Theta})}{Var(RF_t | \mathcal{F}_{t-1}^{\Theta})} + \sum_{j=1}^{M} \nu_j w_j, \tag{4.12}$$

where

$$w_j = \frac{\left(\sum_{s_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(s_t) | \mathcal{F}_{t-1}^{\Theta}]) p(s_t = j | s_{t-1}, \theta) p(s_{1:t-1} | \mathcal{F}_{t-1}, \theta)\right)}{\sum_{j=1}^{M} \left(\sum_{s_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(s_t) | \mathcal{F}_{t-1}^{\Theta}]) p(s_t = j | s_{t-1}, \theta) p(s_{1:t-1} | \mathcal{F}_{t-1}, \theta)\right)},$$

$$E(a(s_t) | \mathcal{F}_{t-1}^{\Theta}) = \sum_{s_{1:t-1}} \sum_{j=1}^{M} a_j p(s_t = j | s_{t-1}, \theta) p(s_{1:t-1} | \mathcal{F}_{t-1}, \theta),$$

$$Cov(\mu(s_t), a(s_t) | \mathcal{F}_{t-1}^{\Theta}) = \sum_{s_{1:t-1}} \sum_{j=1}^{M} \left(\mu_j a_j - \mu_j E[a(s_t) | \mathcal{F}_{t-1}^{\Theta}]\right) p(s_t = j | s_{t-1}, \theta) p(s_{1:t-1} | \mathcal{F}_{t-1}, \theta),$$

$$V(RF_t | \mathcal{F}_{t-1}^{\Theta}) = \sum_{j=1}^{M} \left(\sum_{s_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(s_t) | \mathcal{F}_{t-1}^{\Theta}]) p(s_t = j | s_{t-1}, \theta) p(s_{1:t-1} | \mathcal{F}_{t-1}, \theta) \right),$$

and $\tau_t^2(k) = \kappa_k + \omega_k \xi_{t-1}^2 + \phi_k \tau_{t-1}^2$ for $k = 1,\ldots,M$ and $t = 1,\ldots,T$.

Remark 4.2.4. If s_t, z_t are independent given (s_{t-1}, z_{t-1}) and z_{t-1} (s_{t-1}) does not cause s_t (z_t) one step ahead given s_{t-1} (z_{t-1}) and $\nu(s_t, z_t) = \nu(s_t)$, then the result by

Alizadeh and Nomikos [2004] is obtained i.e.

$$\begin{aligned}
h_t &= \sum_{j=1}^{M} \nu_j \left(\frac{\left(\sum_{s_{1:t-1}} p(s_t = j|s_{t-1}, \theta) p(s_{1:t-1}|\mathcal{F}_{t-1}, \theta)\right)}{\sum_{j=1}^{M} \left(\sum_{s_{1:t-1}} p(s_t = j|s_{t-1}, \theta) p(s_{1:t-1}|\mathcal{F}_{t-1}, \theta)\right)} \right) \\
&= \sum_{j=1}^{M} \nu_j \left(\frac{\left(\sum_{s_{1:t-1}} p(s_t = j, s_{1:t-1}|\mathcal{F}_{t-1}, \theta)\right)}{\sum_{j=1}^{M} \left(\sum_{s_{1:t-1}} p(s_t = j, s_{1:t-1}|\mathcal{F}_{t-1}, \theta)\right)} \right) \\
&= \sum_{j=1}^{M} \nu_j \left(\frac{p(s_t = j|\mathcal{F}_{t-1}, \theta)}{\sum_{j=1}^{M} \left(p(s_t = j|\mathcal{F}_{t-1}, \theta)\right)} \right) \\
h_t &= \sum_{j=1}^{M} \nu_j p(s_t = j|\mathbf{y}_{1:t-1}, \theta).
\end{aligned} \tag{4.13}$$

Furthermore, we expect a better estimate of the optimal hedge ratio using the above outlined framework over constant hedge ratio since the model allows for shifts in the mean and volatility of both RS_t and RF_t and recognizes the relationship between them. As noted in Section 4.2, the model parameters in Equation (4.10) are not known in practice. In this respect, a natural approach to solving this problem will be to apply the plug-in principle. Alternatively, following the Bayesian paradigm outlined above we have the following proposition.

Proposition 4.2.5. *Assume that the observations are generated by the process described in (4.7). Then under certain regularity conditions the Bayesian conditional minimum hedge ratio at time t is the solution to*

$$h_t^{BAY} = \arg\min_{h \in H} \{E(Var(RS_t - hRF_t|\mathcal{F}_{t-1}^{\Theta})|\mathcal{F}_{t-1}))\} \tag{4.14}$$

which is given by

$$h_t^{BAY} = \frac{\int_{\Theta}[Cov(\mu(s_t), a(z_t)|\mathcal{F}_{t-1}^{\Theta})]p(\theta|\mathbf{y}_{1:t-1})d\theta}{\int_{\Theta}[Var(RF_t|\mathcal{F}_{t-1}^{\Theta})]p(\theta|\mathbf{y}_{1:t-1})d\theta} + \sum_{i,j=1}^{M} \int_{\Theta} \nu_{ij} w_{ij}(\theta|\mathbf{y}_{1:t-1})d\theta, \tag{4.15}$$

where,

$$w_{ij}(\theta|\mathbf{y}_{1:t-1}) = \frac{\sum_{(s,z)_{1:t-1}}(a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}_{t-1}^{\Theta}])\pi_{ij,..}p((s,z)_{1:t-1},\theta|\mathbf{y}_{1:t-1})}{\sum_{i,j=1}^{M}\left(\int_{\Theta}\sum_{(s,z)_{1:t-1}}(a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}_{t-1}^{\Theta}])\pi_{ij,..}p((s,z)_{1:t-1},\theta|\mathbf{y}_{1:t-1})d\theta\right)},$$

$$\pi_{ij,..} = p(s_t = i, z_t = j|s_{t-1}, z_{t-1}, \theta),$$

$\tau_t^2(k) = \kappa_k + \omega_k\xi_{t-1}^2 + \phi_k\tau_{t-1}^2$ *for* $k = 1,\ldots,M,\ t = \bar{t},\ldots,T$ *and* $\bar{t}$ *is the minimum number of observations needed for the posterior distribution of* θ *to be proper.*

Proof: See Appendix (4.B).

Similar to Proposition (4.2.1), Proposition (4.2.5) states that the Bayesian optimal hedge ratio at any point in time can be determined by two components. The first component measures the expected covariance between the intercepts divided by the expected variance of the returns on futures after incorporating all available information about the unknown parameters through their joint posterior distribution. Conditional on past observations, the second component is the expected hedge ratio subject to a modified joint posterior distribution of the unknown parameters.

4.2.2 Computational requirement

An important ingredient needed in the Bayesian optimal hedge ratio defined by (4.15) is the posterior distribution of the augmented parameter vector $p((s,z)_{1:t-1},\theta|\mathbf{y}_{1:t-1}),\ t = \bar{t},\ldots,T$. These quantities cannot be identified with any known distribution. This limitation makes the evaluation of (4.15) a non-trivial one. We shall address this problem by using a simulation based technique.

The computation of the MV hedge ratio will be broken down into two main stages. The first part consist of approximating the posterior distribution of the unknown parameters vector given past observations, $p((s,z)_{1:t-1}, \theta|\mathbf{y}_{1:t-1})$, while the second part involves evaluating the hedge ratio.

Following Billio et al. [2012], we describe an efficient simulation based technique for Bayesian approximation of the posterior probability, $p((s,z)_{1:t-1}, \theta|\mathbf{y}_{1:t-1})$. The Bayesian approach is based on MCMC Gibbs algorithm which allows us to circumvent the path dependence problem inherent in MS-GARCH models and efficiently sample the state trajectories. The purpose of this algorithm is to generate samples from the posterior distribution, $p((s,z)_{1:t-1}, \theta|\mathbf{y}_{1:t-1})$, which are then used in the second stage for approximating the moments in (4.15).

We assume fairly informative prior for θ_π and independent uniform prior for θ_u^{RS}, θ_a^{RF}, θ_σ and θ_τ and denote with $f(\theta)$ the joint prior density. To avoid label switching we assume that $\gamma_1 < \gamma_2 < \cdots < \gamma_M$, $\kappa_1 < \kappa_2 < \cdots < \kappa_M$ i.e. identifiability restriction. The posterior density of the augmented parameter vector given by

$$f(\theta, (s,z)_{1:t}|RS_{1:t}, RF_{1:t}) \propto f(RS_{1:t}|(s,z)_{1:t}, \theta, RF_{1:t}) f(RF_{1:t}|(s,z)_{1:t}, \theta) p((s,z)_{1:t}|\theta) f(\theta) \tag{4.16}$$

for $t = \bar{t}, \ldots, T$, cannot be identified with any standard distribution, hence we cannot sample directly from it. In this respect, we apply Gibbs sampling technique whereby for each $t = \bar{t}, \ldots, T$, our Gibbs sampler generate samples by iteratively sampling from the following full conditional distributions:

1. $p((s,z)_{1:t}|\theta, RS_{1:t}, RF_{1:t})$,

2. $f(\theta_\pi|\theta_u^{RF}, \theta_a^{RS}, \theta_\sigma, \theta_\tau, (s,z)_{1:t}, RS_{1:t}, RF_{1:t}) = f(\theta_\pi|(s,z)_{1:t})$, and

3. $f(\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau|\theta_\pi, (s,z)_{1:t}, RS_{1:t}, RF_{1:t}) = f(\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau|(s,z)_{1:t}, RS_{1:t}, RF_{1:t})$.

These full conditional distributions are easier to manage and sample from since they can either be associated with a known distribution or simulated from by a lower dimensional auxiliary sampler.

The full joint distribution of the state variables, $s_{1:t}$, given the parameter values and return series

$$p((s,z)_{1:t}|\theta, RS_{1:t}, RF_{1:t}) \propto f(RS_{1:t}|RF_{1:t}, \theta, (s,z)_{1:t}) f(RF_{1:t}|\theta, (s,z)_{1:t}) \tag{4.17}$$

is a non-standard distribution. In view of this, we consider a Metropolis Hastings (MH) strategy for generating proposals for the state variables. We construct the proposal distribution by first considering an approximation of the regime switching GARCH model and then derive the joint distribution of the state variables. See Billio et al. [2012] for alternative approximations. For expository purpose, we apply the simplified Klaassen's approximation given in Billio et al. [2012]. Samples of the state trajectory are then drawn by Forward Filter Backward sampling scheme.

Let the proposal distribution be denoted by

$$q((s,z)_{1:t}|\theta, RS_{1:t}, RF_{1:t}) = q(s_t, z_t|\theta, RS_{1:t}, RF_{1:t}) \prod_{r=1}^{t-1} q(s_r, z_r|s_{r+1}, z_{r+1}, \theta, RS_{1:r}, RF_{1:r}), \tag{4.18}$$

for $t = \bar{t}, \ldots, T$ and where

$$\begin{aligned} &q(s_r, z_r|\theta, RS_{1:r-1}, RF_{1:r-1}) \\ &= \sum_{i,j=1}^{M} p(s_r, z_r|s_{r-1}=i, z_{r-1}=j) q(s_{r-1}=i, z_{r-1}=j|\theta, RS_{1:r-1}, RF_{1:r-1}), \end{aligned}$$

$$\begin{aligned} &q(s_r, z_r|s_{r+1}, z_{r+1}, \theta, RS_{1:r}, RF_{1:r}) \\ &= \frac{p(s_{r+1}, z_{r+1}|s_r, z_r) q(s_r, z_r|\theta, RS_{1:r}, RF_{1:r})}{\sum_{i,j=1}^{M} p(s_{r+1}, z_{r+1}|s_r=i, z_r=j) q(s_r=i, z_r=j|\theta, RS_{1:r}, RF_{1:r})}, \end{aligned}$$

$$\begin{aligned} &q(s_r, z_r|\theta, RS_{1:r}, RF_{1:r}) \\ &= \frac{g(RS_r|s_r, z_r, \theta, RS_{1:r-1}, RF_{1:r}) g(RF_r|s_r, z_r, \theta, RF_{1:r-1}) q(s_r, z_r|\theta, RS_{1:r-1}, RF_{1:r-1})}{g(RS_{1:r}, RF_{1:r}|\theta)}, \end{aligned}$$

$$g(RS_r|s_r, z_r, \theta, RS_{1:r-1}, RF_{1:r}) \propto \prod_{r^*=1}^{r} \frac{1}{\sigma_{(RS)r^*}} \exp\left(\frac{-(RS_{r^*} - \mu(s_{r^*}) - \nu(s_{r^*}, z_{r^*})RF_{r^*})^2}{2\sigma^2_{(RS)r^*}}\right),$$

$$g(RF_r|s_r, z_r, \theta, RF_{1:r-1}) \propto \prod_{r^*=1}^{r} \frac{1}{\tau_{(RF)r^*}} \exp\left(-\frac{(RF_{r^*} - a(z_{r^*}))^2}{2\tau^2_{(RF)r^*}}\right),$$

$$\sigma^2_{(RS)r^*} = \gamma(s_{r^*}) + \alpha(s_{r^*})\left(RS_{r^*-1} - \sum_{i,j=1}^{M}(\mu_i + \nu_{ij}RF_{r^*-1})q(s_{r^*-1} = i, z_{r^*-1} = j|RS_{1:r^*-1}, RF_{1:r^*-1})\right)$$
$$+ \beta(s_{r^*})\left(\sum_{i,j=1}^{M} \sigma^2_{(RS)r^*-1}(i)q(s_{r^*-1} = i, z_{r^*-1} = j|RS_{1:r^*-1}, RF_{1:r^*-1})\right),$$

$$\tau^2_{(RF)r^*} = \kappa(z_{r^*}) + \omega(z_{r^*})\left(RF_{r^*-1} - \sum_{m=1}^{M}(a_m q(z_{r^*-1} = m|RS_{1:r^*-1}, RF_{1:r^*-1}))\right)^2$$
$$+ \psi(s_{r^*})\left(\sum_{m=1}^{M} \tau^2_{(RF)r^*-1}(m)q(z_{r^*-1} = m|RS_{1:r^*-1}, RF_{1:r^*-1})\right),$$

for $i = 1, \ldots, t$. The full conditional of θ_π is Dirichlet under Dirichlet prior distribution assumption and the posterior density of $(\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau)$

$$f(\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau|(s,z)_{1:t}, RS_{1:t}, RF_{1:t})$$
$$\propto \prod_{r=1}^{t} \frac{1}{\sigma_r} \exp\left(-\frac{(RS_r - \mu(s_r) - \nu(s_r)RF_r)^2}{2\sigma_r^2}\right) \prod_{r=1}^{t} \frac{1}{\tau_r} \exp\left(\frac{-(RF_r - a(z_r))^2}{2\tau_r^2}\right) \tag{4.19}$$

is non-standard. Hence, we apply adaptive Metropolis-Hastings (MH) sampling technique for this step of the Gibbs algorithm. A summary of this step is summarized in the Algorithm 1.

Algorithm 1: Posterior approximation

For each $t = \bar{t}, \ldots, T$

1. Choose a starting value $(s^{(0)}, z^{(0)})_{1:t}$ and $\theta^{(0)}$.

2. Let $(s^{(g-1)}, z^{(g-1)})_{1:t}$, $\theta^{(g-1)}$ and $p^{(g-1)}(s_t|\mathcal{F}_{t-1}, \theta)$ respectively be the state trajectory, parameter set and prediction probability at $(g-1)$th iteration.

3. Draw $(s, z)_{1:t}$ using FFBS from $q((s, z)_{1:t}|\theta, RS_{1:t}, RF_{1:t})$ and identify $q(s_t, z_t|\mathcal{F}_{t-1}, \theta)$ from the forward filter.

4. Draw $u \sim \mathcal{U}_{[0,1]}$ and set

$$(s^{(g)}, z^{(g)})_{1:t} = \begin{cases} (s, z)_{1:t} & \text{if } u \leq \alpha((s, z)_{1:t}, (s^{(g-1)}, z^{(g-1)})_{1:t}), \\ (s^{(g-1)}, z^{(g-1)})_{1:t} & \text{otherwise,} \end{cases}$$

where,

$$\alpha((s, z)_{1:t}, (s^{(g-1)}, z^{(g-1)})_{1:t}) = \left(1, \frac{p((s, z)_{1:t}|\theta, RS_{1:t}, RF_{1:t})q((s^{(g-1)}, z^{(g-1)})_{1:t}|\theta, RS_{1:t}, RF_{1:t})}{q((s, z)_{1:t}|\theta, RS_{1:t}, RF_{1:t})p((s^{(g-1)}, z^{(g-1)})_{1:t}|\theta, RS_{1:t}, RF_{1:t})}\right).$$

5. Draw θ_π from a Dirichlet distribution.

6. Draw $\theta_{-\pi}$ from $g(\theta_{-\pi}|(s^{(g-1)}, z^{(g-1)})_{1:t}, RS_{1:t}, RF_{1:t})$.

7. Draw $u \sim \mathcal{U}_{[0,1]}$ and set

$$\theta_{-\pi}^{(g)} = \begin{cases} \theta_{-\pi} & \text{if } u \leq \alpha(\theta_{-\pi}, \theta_{-\pi}^{(g-1)}), \\ \theta_{-\pi}^{(g-1)} & \text{otherwise,} \end{cases}$$

where

$$\alpha(\theta_{-\pi}, \theta_{-\pi}^{(g-1)}) = \left(1, \frac{f(\theta_{-\pi}|(s, z)_{1:t}, RS_{1:t}, RF_{1:t})g(\theta_{-\pi}^{(g-1)}|(s, z)_{1:t}, RS_{1:t}, RF_{1:t})}{g(\theta_{-\pi}|(s, z)_{1:t}, RS_{1:t}, RF_{1:t})f(\theta_{-\pi}^{(g-1)}|(s, z)_{1:t}, RS_{1:t}, RF_{1:t})}\right)$$

In the second stage, samples generated from the posterior distribution, $p((s,z)_{1:t-1},\theta|\mathcal{F}_{t-1})$, will be used for computing the Bayesian hedge ratio 4.15. Let G represent the number of Monte Carlo samples from $p((s,z)_{1:t-1},\theta|\mathcal{F}_{t-1})$, then

$$\hat{h}_t^{BAY} = \frac{\sum_{g=1}^{G}[Cov(\mu^{(g)}(s_t), a^{(g)}(z_t)|\mathcal{F}_{t-1}^{\Theta})]}{\sum_{g=1}^{G}[Var^{(g)}(RF_t|\mathcal{F}_{t-1})]} + \frac{1}{G}\sum_{i,j=1}^{M}\sum_{g=1}^{G}\nu_{ij}^{(g)} w_{ij}(\theta^{(g)}|\mathbf{y}_{1:t-1})$$

$$w_{ij}(\theta^{(g)}|\mathbf{y}_{1:t-1})$$
$$= \frac{((\tau_t^{(g)})^2(j) + (a_j^{(g)})^2 - a_j^{(g)}E[a^{(g)}(z_t)|\mathcal{F}_{t-1}^{\Theta}])p^{(g)}(s_t=i, z_t=j|s_{t-1}^{(g)}, z_{t-1}^{(g)}, \theta^{(g)})}{\frac{1}{G}\sum_{j,k=1}^{M}\sum_{g=1}^{G}(((\tau_t^{(g)})^2(k) + (a_k^{(g)})^2 - a_k^{(g)}E[a^{(g)}(z_t)|\mathcal{F}_{t-1}^{\Theta}])p^{(g)}(s_t=j, z_t=k|s_{t-1}^{(g)}, z_{t-1}^{(g)}, \theta^{(g)}))}.$$

We summarize in Algorithm 2 the steps involved in this stage.

Algorithm 2: Hedging

For each $t = \bar{t}, \ldots, T$

1. Compute the moments and substitute into 4.2.2

$$E[a^{(g)}(s_t)|\mathcal{F}^{\Theta}_{t-1}]$$
$$= \sum_{i=1}^{M}\sum_{j=1}^{M} a_j^{(g)} p^{(g)}(s_t = i, z_t = j|s_{t-1}^{(g)}, z_{t-1}^{(g)}, \theta^{(g)})$$

$$Cov(\mu^{(g)}(s_t), a^{(g)}(z_t)|\mathcal{F}^{\Theta}_{t-1})$$
$$= \sum_{i=1}^{M}\sum_{j=1}^{M} \left(\mu_i^{(g)} a_j^{(g)} - \mu_i^{(g)} E[a^{(g)}(z_t)|\mathcal{F}_{t-1}]\right) p^{(g)}(s_t = i, z_t = j|s_{t-1}^{(g)}, z_{t-1}^{(g)}, \theta^{(g)})$$

$$Cov(\nu^{(g)}(s_t)RF_t, RF_t|\mathcal{F}^{\Theta}_{t-1})$$
$$= \sum_{i=1}^{M}\sum_{j=1}^{M} \nu_{ij}^{(g)}((a_j^{(g)})^2 + (\tau_t^{(g)})^2(j) - a_j^{(g)} E[a^{(g)}(z_t)|\mathcal{F}_{t-1}])p^{(g)}(s_t = i, z_t = j|s_{t-1}^{(g)}, z_{t-1}^{(g)}, \theta^{(g)})$$

$$Var(RF_t|\mathcal{F}^{\Theta}_{t-1})$$
$$= \sum_{i=1}^{M}\sum_{j=1}^{M}((a_j^{(g)})^2 + (\tau_t^{(g)})^2(j) - a_j^{(g)} E[a^{(g)}(z_t)|\mathcal{F}_{t-1}])p^{(g)}(s_t = i, z_t = j|s_{t-1}^{(g)}, z_{t-1}^{(g)}, \theta^{(g)})$$

It is worth noting that the decision problem characterized by Proposition 4.2.5 may be classified as a sequential estimation problem. This is because, in contrast to a fixed decision problem, as new observation $\mathbf{y}_{t+1}$ arrives, the hedger updates the posterior distribution, $f(\theta|\mathbf{y}_{1:t})$, about the unknown parameters and by induction revises the hedge ratio. In our computational procedure, at each date t MCMC algorithm is employed for drawing samples from the posterior probability distribution of the unknown parameters which are then used in computing the moments in the Bayesian hedge ratio. A drawback of our Bayesian estimation approach is the potential computational burden involved running the MCMC algorithm on the posterior probability distribution at each date. However, it can be argued that, the procedure remain feasible in practice since the computation of hedging ratio may be done overnight and result will be available at the beginning of the next trading day. Alternative procedures such as the sequential MCMC proposed by Yang and Dunson [2013] or sequential monte carlo (see Doucet et al. [2001]) may be used to reduce the computing time when a timely updating of the hedge ratio is required.

4.3 Empirical Application

To illustrate the proposed method, we use daily closing energy prices for West Texas Intermediate (WTI) crude oil futures for the period September 14, 2001 to July 31, 2013 (2967 observations). Both spot and futures daily settlement prices were obtained from the US Energy information Agency (http://www.eia.doe.gov). The daily returns are computed using the first difference of the natural logarithm of the daily settlements. Figure (4.1) displays the sample path of crude oil squared returns on spot and futures. We observe volatility clustering, which calls for the use of MS-GARCH models. We consider a two regime ($M = 2$) MS-GARCH model with

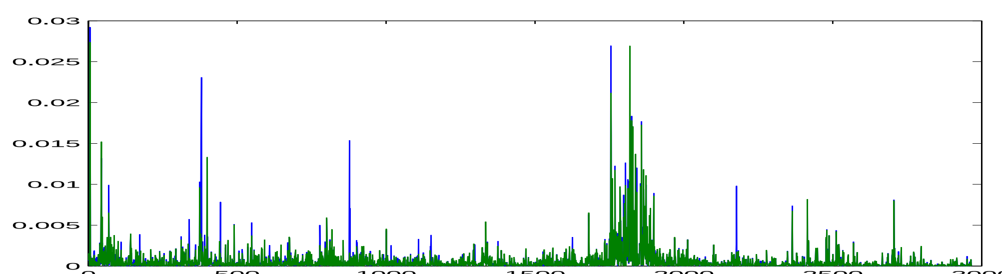


Figure 4.1: Graphs for daily squared returns on WTI crude oil spot and futures from September 14/09/2001 to 31/07/2013

($a(z_t) = 0$ and $\mu(s_t) = 0$) since the sample mean of the both returns on futures and spot are close to zero. Before we proceed with estimating the hedge ratio, we consider a full sample estimation of the parameters of the MS-GARCH model under different assumptions on the hidden state process (ie. independent, dependent and same state variables). This estimation will help us investigate the potential impact on the hedged portfolio at time t if the state of the futures returns at time $t-1$ remains the same or changes and vice versa.

4.3.1 Full sample Estimation

In this paper, we perform 10000 Gibbs iterations after convergence according to the Geweke's diagnostic. To reduce serial correlation of the draws, we consider every 10th draw after convergence of the Gibbs iteration in the results presented below. Table 4.1 to 4.4 show the estimation results for two state models using the full sample of observation described above. From the estimated parameters in Table 4.1, regime 1 may be labeled as the low volatility state. Except for the constrained multichain MS-GARCH model, we observe from Table 4.1 that the volatility persistence of the hedged portfolio measured by the sum of garch parameter, β, and the arch parameter, α, is higher in regime 1 than in regime 2. In other words, large persistence tends to be associated with low volatility regime. This observation may largely be a reflection of the dependence assumption between the chains driving the two series as observed in the unconstrained multichain MS-GARCH model and the single chain MS-GARCH model case.

The transition probabilities in Table 4.2 shows that the probability that the hedged portfolio and futures return simultaneously remain in the high regime is very low. Whereas, from Table 4.4 we observe that the single chain MS-GARCH model gives a relatively high probability for the two variables to be in the high state simultaneously. The implication of this observation is that, when possible misalignment between the states of the chains driving the two dynamics are not taken into account, our results may be a reflection of an under or over estimation exercise. Nevertheless, a consistent deduction from all the MS-GARCH specifications under consideration is that when both returns are in the low regime at time $t-1$, it is probable to maintain this scenario in the next period. Also, Table 4.2 suggests that when the returns on the hedged portfolio is in a different state with respect to the returns on the futures at time $t-1$, then the most probable scenario at time t will be the alignment of the futures to the same scenario of the hedged portfolio.

Table 4.1: Parameter estimate of the MSGRACH model and standard deviation in parenthesis.

	MC-f-MSGARCH	MC-c-MSGARCH	SC-MSGARCH
ν_{11}	0.994(0.0011)	0.993(0.0014)	0.991(0.0013)
ν_{12}	0.629(0.0097)		
ν_{21}	0.947(0.0011)	0.875(0.0128)	0.829(0.0189)
ν_{22}	0.055(0.0097)		
γ_1	1.23e-06(4.93e-08)	1.62e-06(7.85e-08)	1.64e-06(2.15e-07)
γ_2	8.33e-05(6.06e-06)	1.14e-04(9.67e-06)	1.65e-04(1.79e-05)
α_1	0.560(0.0369)	0.363(0.0310)	0.868(0.0501)
α_2	0.586(0.0554)	0.632(0.0708)	0.091(0.0503)
β_1	0.037(0.0022)	0.005(0.0032)	0.022(0.0086)
β_2	0.292(0.0525)	0.325(0.0675)	0.442(0.0873)
κ_1	9.76e-06(3.25e-06)	1.11e-06(8.99e-07)	7.14e-06(3.315e-06)
κ_2	9.70e-05(4.37e-05)	5.48e-05(1.34e-05)	4.73e-05(1.813e-05)
ω_1	0.073(0.0104)	0.026(0.0062)	0.062(0.0122)
ω_2	0.093(0.0144)	0.122(0.0220)	0.084(0.0226)
ψ_1	0.908(0.0123)	0.965(0.0067)	0.918(0.0176)
ψ_2	0.794(0.0097)	0.789(0.0388)	0.872(0.0370)

Notes: SC-MSGARCH stands for single chain MS-GARCH; MC-c-MSGARCH stands for constrained Multichain MS-GARCH model; and MC-f-MSGARCH stands for unconstrained Multichain MS-GARCH

Table 4.2: Transition matrix for MC-f-MSGARCH model.

	$s_{t-1}=1, z_{t-1}=1$	$s_{t-1}=1, z_{t-1}=2$	$s_{t-1}=2, z_{t-1}=1$	$s_{t-1}=2, z_{t-1}=2$
$s_t=1, z_t=1$	0.9124	0.6383	0.2672	0.2176
$s_t=1, z_t=2$	0.0026	0.0866	0.0168	0.2635
$s_t=2, z_t=1$	0.0766	0.0534	0.6682	0.3781
$s_t=2, z_t=2$	0.0084	0.2217	0.0478	0.1408

Table 4.3: Transition matrix for MC-c-MSGARCH model.

(a) Hedged portfolio

	$s_{t-1} = 1$	$s_{t-1} = 2$
$s_t = 1$	0.894	0.383
$s_t = 2$	0.106	0.617

(b) Futures

	$z_{t-1} = 1$	$z_{t-1} = 2$
$z_t = 1$	0.974	0.059
$z_t = 2$	0.026	0.941

Table 4.4: Transition matrix for SC-MSGARCH model.

	$s_{t-1} = 1$	$s_{t-1} = 2$
$s_t = 1$	0.930	0.424
$s_t = 2$	0.070	0.576

Also of interest in the unconstrained multichain MS-GARCH framework is the ability to say in clear terms the effect of the possibility of changing the regime for a single series, given a certain scenario in the previous period i.e.

$$p(s_t = i | s_{t-1} = h, z_{t-1} = r) \quad \text{and} \quad p(z_t = j | s_{t-1} = h, z_{t-1} = r) \quad \forall \;\; i, j, h, r = 1, 2$$

In Table 4.5, we report these probabilities. The influence of z_{t-1} on the changes in regime for the hedged portfolio are evident; in fact, the probability of hedged portfolio staying in regime 1, when the futures was in regime 1 in the previous month, is 0.92, but decreases to 0.72 when the futures was in regime 2. In a similar way, the futures remains in regime 1 with an 98% chance when the hedged portfolio was in the same regime, but switches to regime 2 with a probability equal to 31% when the hedged portfolio was in regime 2.

Table 4.5: Conditional probabilities for the MC-f-MSGARCH model.

	$s_{t-1} = 1, z_{t-1} = 1$	$s_{t-1} = 1, z_{t-1} = 2$	$s_{t-1} = 2, z_{t-1} = 1$	$s_{t-1} = 2, z_{t-1} = 2$
$s_t = 1$	0.9150	0.7249	0.2840	0.4811
$s_t = 2$	0.0850	0.2751	0.7160	0.5189
$z_t = 1$	0.9890	0.6917	0.9354	0.5957
$z_t = 2$	0.0110	0.3083	0.0646	0.4043

Lastly, the correlations between the returns on the spot and the futures can be obtained by evaluating

$$\rho_t = \frac{1}{\sqrt{\frac{\sigma_t^2}{\nu_t^2 \tau_t^2} + 1}}.$$

Unlike the single chain MS-GARCH model with two correlations, the multichain MS-GARCH models have four possible correlation regime at each point in time. To have an idea of the relative importance of the correlations in each MS-GARCH specifications, we replace the time varying variance with their respective regime unconditional variance. In the unconstrained (constrained) multichain case, when both spot and futures are in the high volatility regime, the correlation is equal to 0.997 (0.989) and when both of them are in the low volatility regime, the correlation is equal to 0.016 (0.389). When spot return is in the high volatility regime and futures return is in the low volatility regime, the correlation is equal to 0.995 (0.998) and when spot return is in the low volatility regime and futures return is in the high volatility regime, the correlation is equal to 0.634 (0.185). In summary, we find that the correlation of spot and futures return series tend to be higher when the spot is in the low volatility regime. The estimated correlations for the first and second regime of the single chain MS-GARCH model are respectively equal to 0.979 and 0.822. This values are somewhere in between the highest correlation and the lowest correlation estimated from the multichain MS-GARCH models. Overall, more model flexibility may be achieved with the unconstrained MS-GARCH model since it has widest correlation range.

Based on the discussion above, it can be deduced that multichain MS-GARCH models have an important role to play in the optimal hedge ratio theory.

4.3.2 Hedge ratio

In order to check whether our proposed model is of practical use, we conduct a sequential estimation exercise to investigate the performance of our proposed model.

For each hedging model, an out-of-sample analysis of hedging performance with daily re-balancing is carried out. On daily basis, an estimate of the MV hedge ratio is obtained and the futures position to be taken at the end of that day until the following day is also determined. The sample is then extended by one day, the hedge ratios re-estimated, and the hedge rebalanced and held until the end of the next day. For each MS-GARCH specification we consider the sequential estimation of the hedge ratio for three sub-periods i.e. 08/08/2006 to 03/01/2007, 01/10/2008 to 25/03/2009 and 15/02/2013 to 31/07/2013. These periods respectively correspond to the period before, during and after the 2008/2009 global financial crisis.

Figure ?? - ?? shows how different regime specific hedge ratios, conditional variance and prediction probabilities have evolved over the three sub-sample periods. For the unconstrained MS-GARCH model, we observe a small range of values for the regime specific hedge ratios (see 4.2(a)) and high variability in the conditional variance of the futures (see 4.3(a)) prior the 2008/2009 global financial crisis. However, after the 2008/2009 global financial crisis, we observe a clear separation of the hedge ratios (4.3(e)) into two groups determined the change in the hidden process on the futures returns. Also, the variability of the conditional variance of the futures is very low (4.3(e)) relative to our observation prior the global financial crisis. Although, the reason for this is not clear, one possible argument is that investors are more careful and are learning from the experience during financial crisis. The application of the single chain MS-GARCH model tend to suffer from an under or over estimation problem arising from the use of a single chain (see 4.4(a)-4.5(e)). In the case of the constrained MS-GARCH model, there no significant difference in the evolution of the hedge ratio before and after the global financial crisis. A direct comparison between the hedge ratios formed by the unconstrained and constrained MS-GARCH suggests that the unconstrained models is more flexible as it produces a wider range of values for the hedge ratio. Overall, the unconstrained MS-GARCH model seem to perform best among the set of models under consideration. Also, the high probability of staying in the low volatility regime implies low transaction cost because the investor only needs to re-balance his portfolio occasionally.

In the above exercise, it is assumed that the prevailing state of the world is known. However, in reality and most certainly, the current state of the world cannot be correctly identified by the hedger. In this situation, the average hedge ratio is implemented and it is computed as using the following:

$$h_t = E_\theta E[\nu(s_t, z_t)|\mathcal{F}_{t-1}] = E_\theta \left[\sum_{m,m'=1}^{M} \nu_m p(s_t = m, z_t = m'|\mathcal{F}_{t-1}) \right], \tag{4.20}$$

and estimated using,

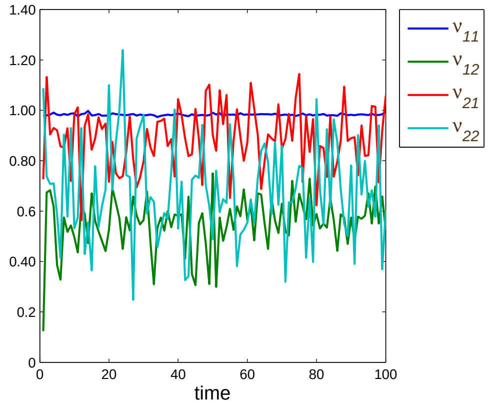
$$h_t = \frac{1}{G} \sum_{i=1}^{G} \sum_{m,m'=1}^{M} \nu_{m,m'}^{(i)} p^{(i)}(s_t = m, z_t = m'|\mathcal{F}_{t-1}), \tag{4.21}$$

where G is the number of Gibbs samples.

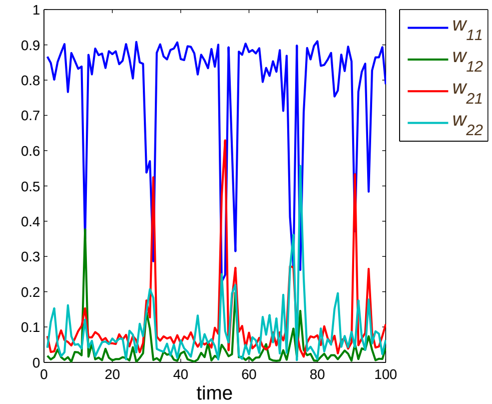
In Figure 4.7, we report the estimation results for each model and compare them with the OLS hedge ratio over the three subsamples. The MS-GARCH hedge ratios display similar time varying characteristics. However, occasionally, we observe that the time varying hedge ratios fall below the OLS hedge ratio. Also, the hedge ratios are observe to shift closer 1 after the global financial crisis. This observation further confirm our earlier deduction that hedgers seem to be more careful in their investment decisions after the global financial crisis.

4.3.3 Hedging effectiveness

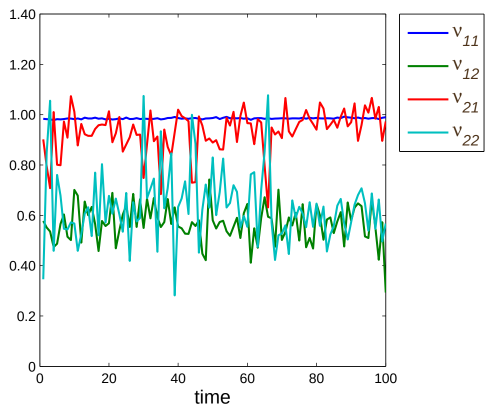
Following the estimation of the hedge ratios, we formally assess the performance of these hedges by first constructing the portfolio implied by the computed hedge ratios daily. Then we calculate the variance of the returns of these portfolios over each



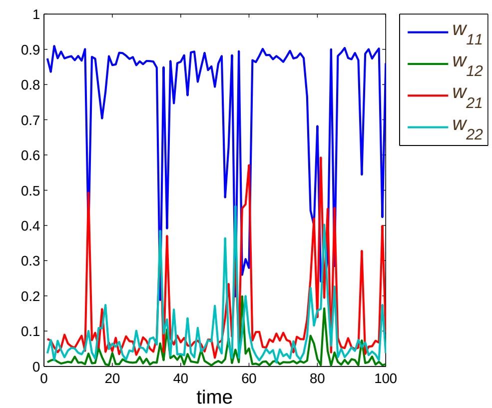
(a) Regime specific hedge ratio before the crisis



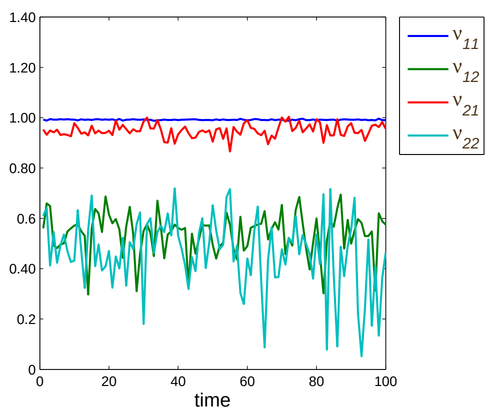
(b) Regime specific weights before the crisis



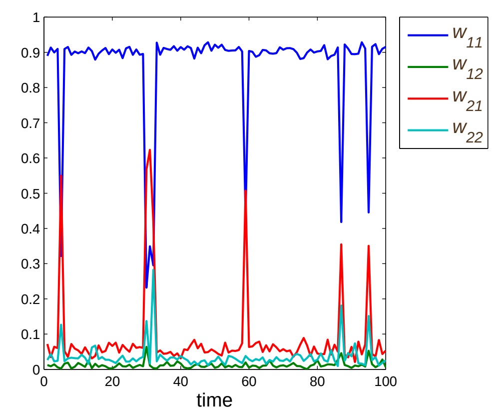
(c) Regime specific hedge ratio during the crisis



(d) Regime specific weights during the crisis

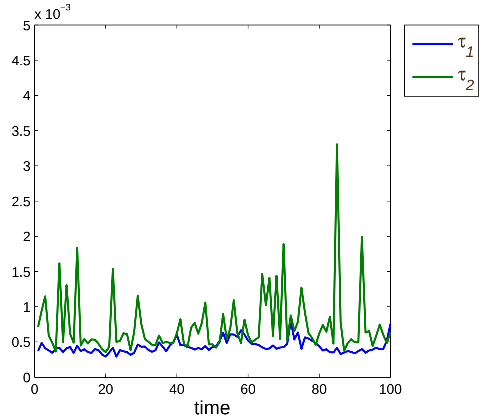


(e) Regime specific hedge ratio after the crisis

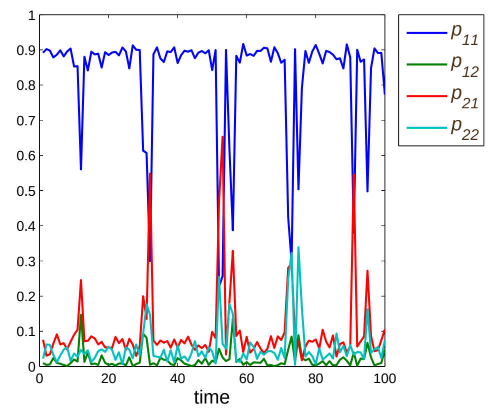


(f) Regime specific weights after the crisis

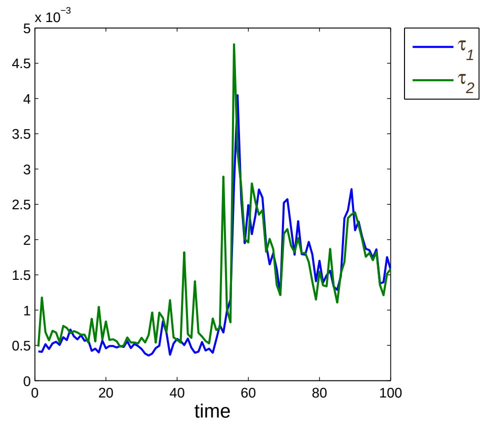
Figure 4.2: Regime specific hedge ratio and corresponding weights for the unconstrained multichain MS-GARCH model (MC-f-MSGARCH). first row 08/08/2006 to 03/01/2007; second row 01/10/2008 to 25/03/2009; third row and 15/02/2013 to 31/07/2013.



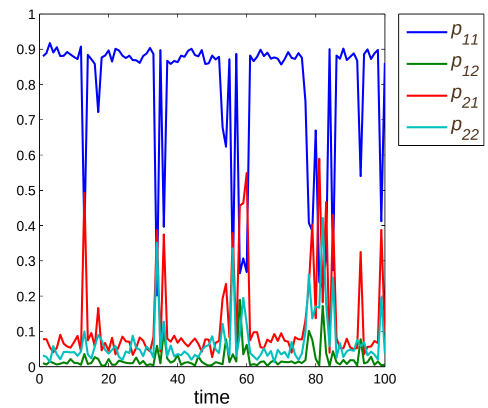
(a) Regime specific conditional variance before the crisis



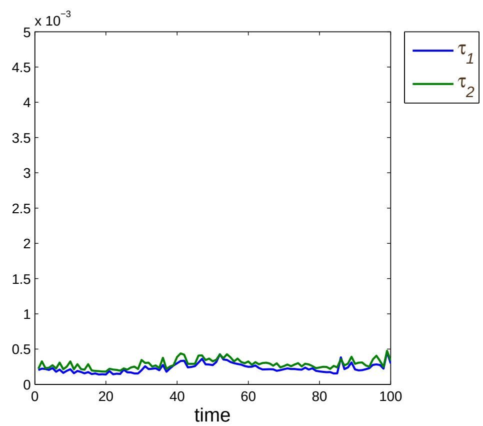
(b) Regime specific prediction probability before the crisis



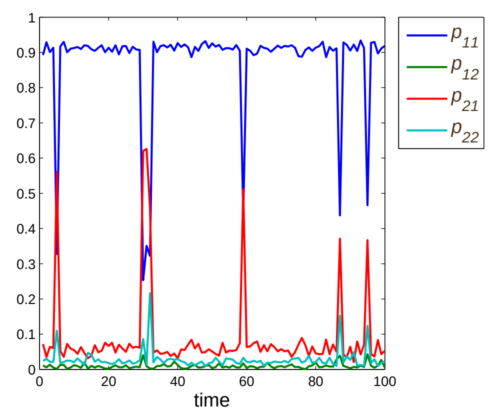
(c) Regime specific conditional variance during the crisis



(d) Regime specific prediction probability during the crisis

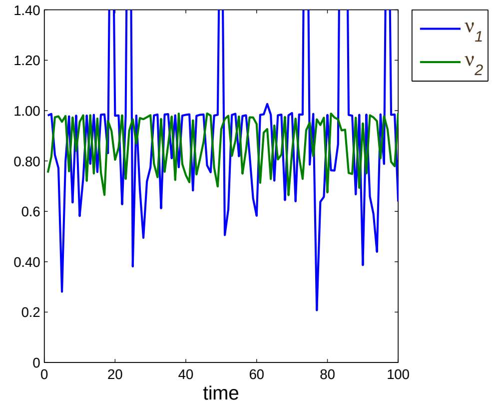


(e) Regime specific conditional variance after the crisis

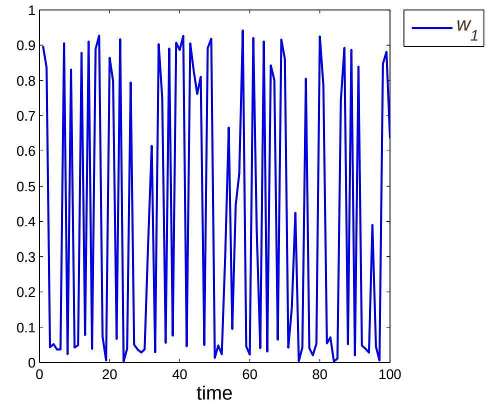


(f) Regime specific prediction probability after the crisis

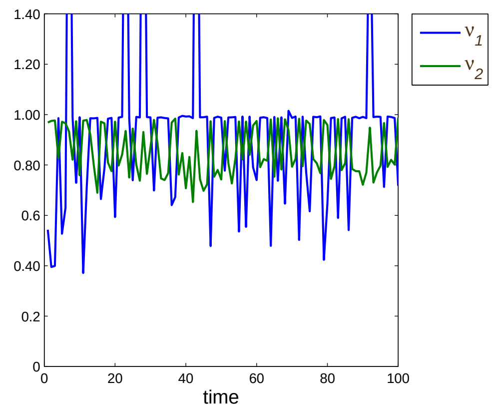
Figure 4.3: Regime specific conditional variance and corresponding predicted probabilities for the unconstrained multichain MS-GARCH model (MC-f-MSGARCH). first row 08/08/2006 to 03/01/2007; second row 01/10/2008 to 25/03/2009; third row and 15/02/2013 to 31/07/2013.



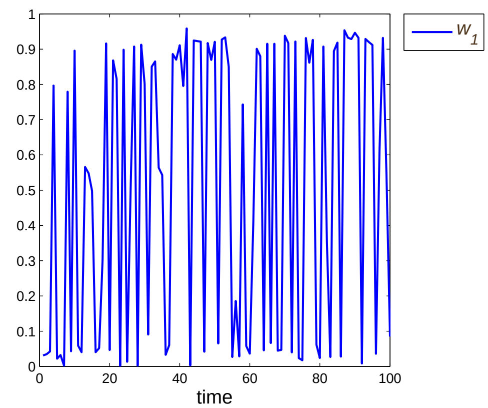
(a) Regime specific hedge ratio before the crisis



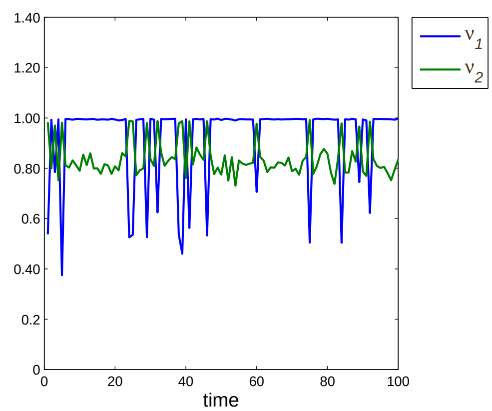
(b) Regime specific weights before the crisis



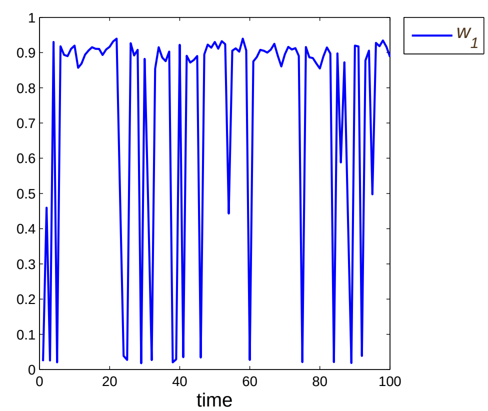
(c) Regime specific hedge ratio during the crisis



(d) Regime specific weights during the crisis

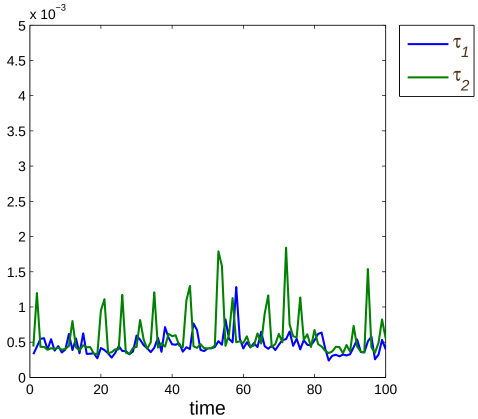


(e) Regime specific hedge ratio after the crisis

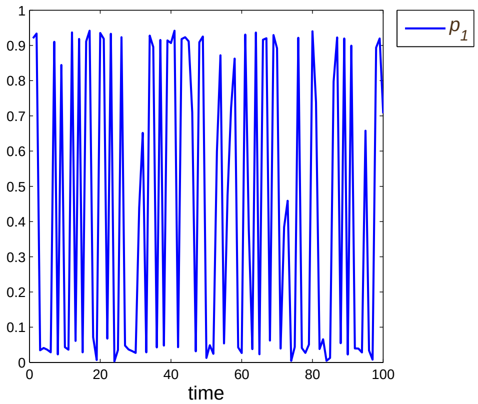


(f) Regime specific weights after the crisis

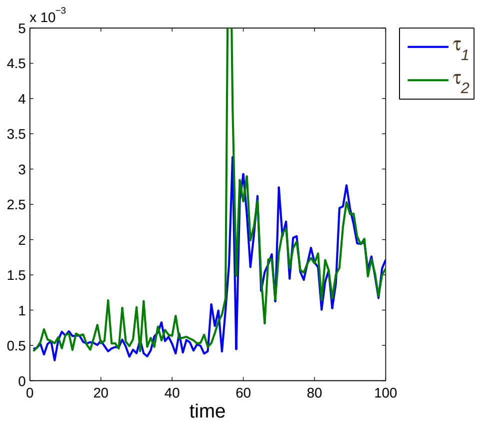
Figure 4.4: Regime specific hedge ratio and corresponding weights for the single chain MS-GARCH model (SC-MSGARCH). first row 08/08/2006 to 03/01/2007; second row 01/10/2008 to 25/03/2009; third row and 15/02/2013 to 31/07/2013.



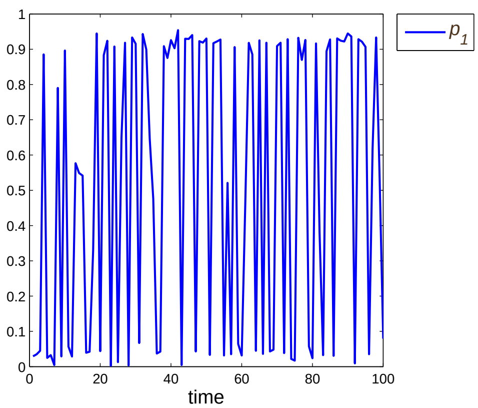
(a) Regime specific conditional variance before the crisis



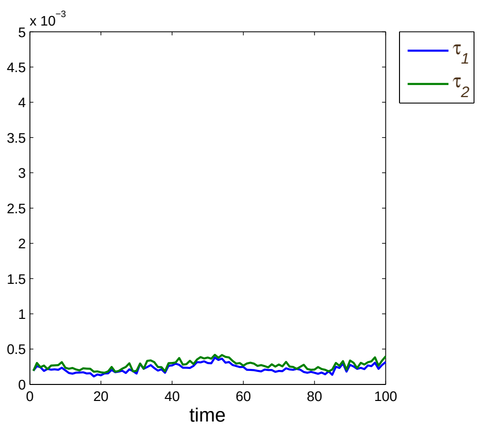
(b) Regime specific prediction probability before the crisis



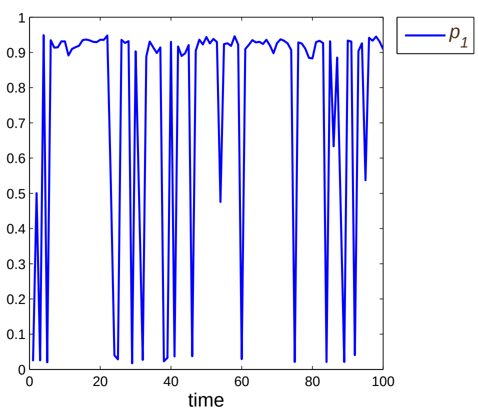
(c) Regime specific conditional variance during the crisis



(d) Regime specific prediction probability during the crisis

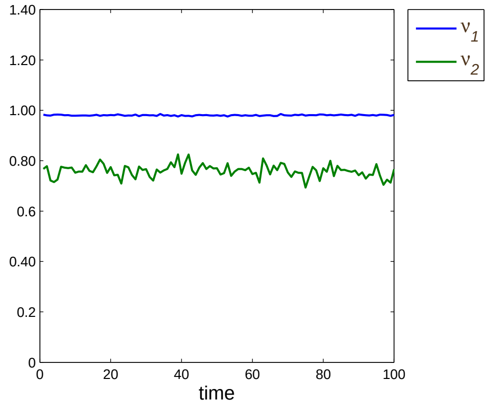


(e) Regime specific conditional variance after the crisis

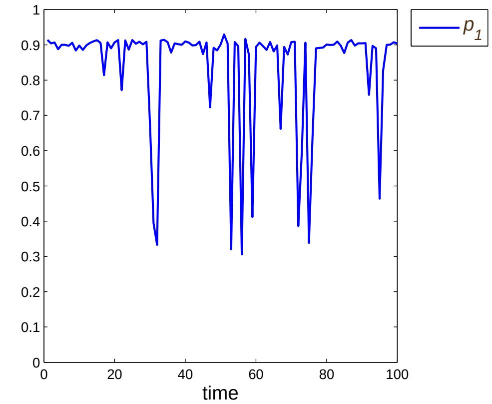


(f) Regime specific prediction probability after the crisis

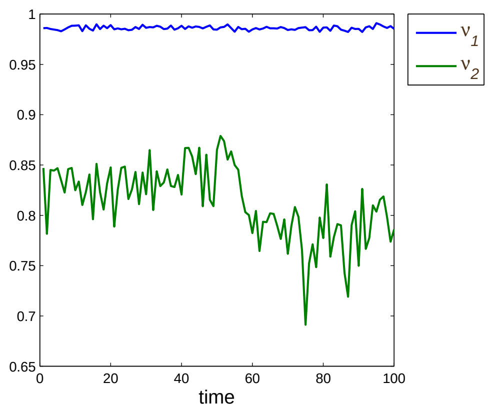
Figure 4.5: Regime specific conditional variance and the corresponding prediction probabilities for the single chain MS-GARCH model (SC-MSGARCH). first row 08/08/2006 to 03/01/2007; second row 01/10/2008 to 25/03/2009; third row and 15/02/2013 to 31/07/2013.



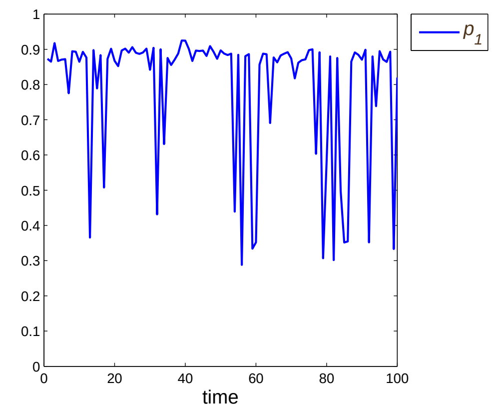
(a) Regime specific hedge ratio before the crisis



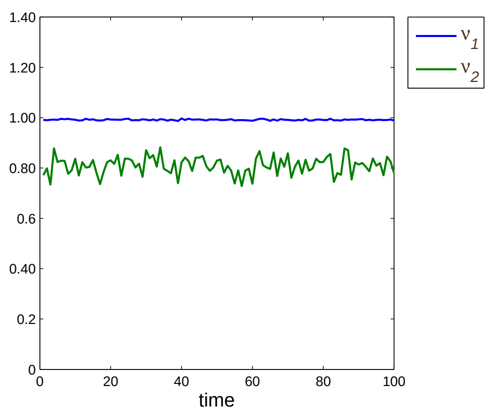
(b) Regime specific weights before the crisis



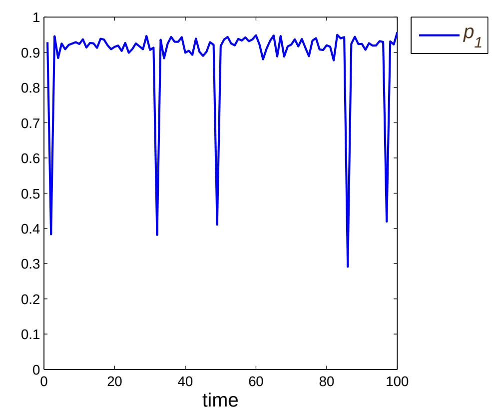
(c) Regime specific hedge ratio during the crisis



(d) Regime specific weights during the crisis

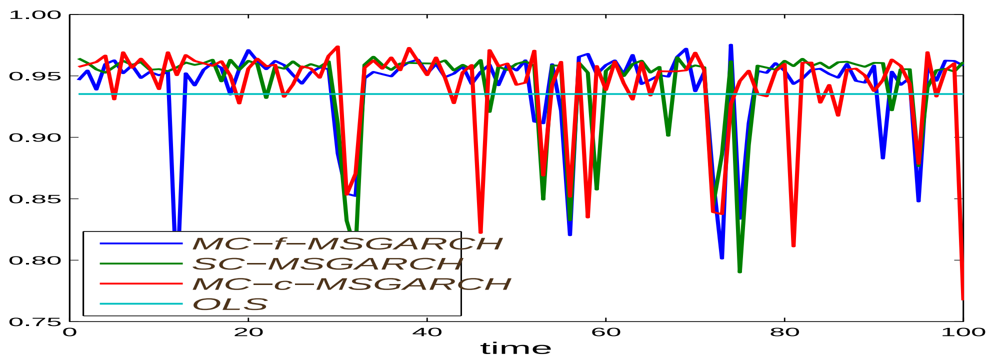


(e) Regime specific hedge ratio after the crisis

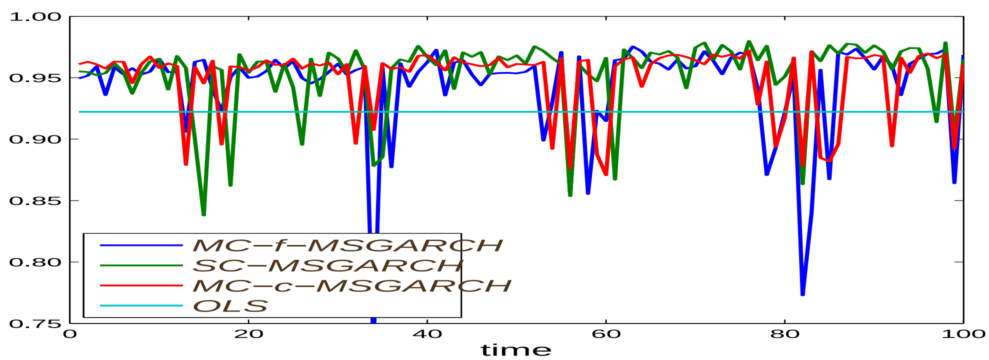


(f) Regime specific weights after the crisis

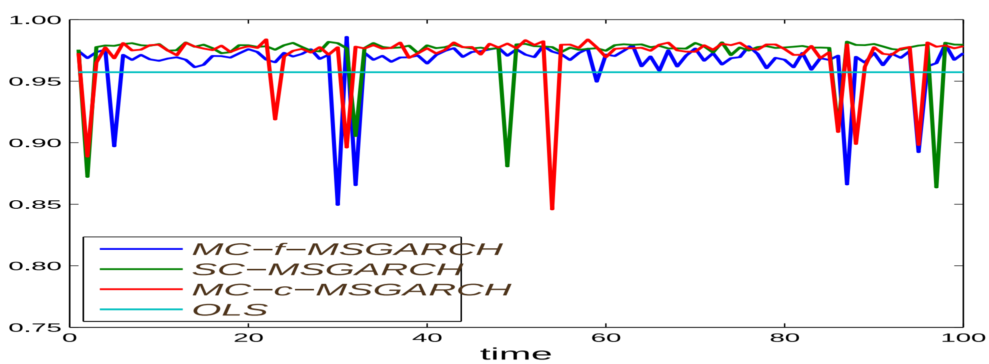
Figure 4.6: Regime specific hedge ratio and corresponding prediction probabilities for the constrained multichain MS-GARCH model (MC-c-MSGARCH). first row 08/08/2006 to 03/01/2007; second row 01/10/2008 to 25/03/2009; third row and 15/02/2013 to 31/07/2013.



(a) Hedge ratio before financial crisis



(b) Hedge ratio during the financial crisis



(c) Hedge ratio after the financial crisis

Figure 4.7: Comparison of average hedge ratio for MC-f-MSGARCH, MC-c-MSGARCH and SC-MSGARCH. first row 08/08/2006 to 03/01/2007; second row 01/10/2008 to 25/03/2009; third row and 15/02/2013 to 31/07/2013.

subsample period. In mathematical forms, we evaluate

$$Var(RS_t - h_t^* RF_t) \tag{4.22}$$

where h_t^* are the estimated hedge ratios. The percentage incremental variance improvement of the MS-GARCH model against the OLS model is calculated as follows

$$\frac{Var(\text{OLS}) - Var(\text{MS-GARCH})}{Var(\text{OLS})} \times 100, \tag{4.23}$$

where Var(OLS) and Var(MS-GARCH) are respectively the variance of the returns on the hedged portfolio (Equation (4.22)) estimated using hedge ratios obtained from the OLS and MS-GARCH models. A positive value of 4.23 is an indication that the MS-GARCH hedge ratio performs better than the OLS hedge ratio. Three different measures of the hedge ratio h_t^* in 4.23 are considered. The first the average hedge ratio given by 4.21, the second is the average hedge ratio at time t given the most probable state at time $t-1$, and the third measure assumes the most probable hedge ratio at time t given the most state of the market at time t.

Table 4.6: Hedging Effectiveness of MS-GARCH against Constant Hedge ratio.

	$h_t^* = E[\nu(s_t, z_t)]$			$h_t^* = E[\nu_t \vert \hat{s}_{t-1}, \hat{z}_{t-1}]$			$h_t^* = \nu(\hat{s}_t.\hat{z}_t)$		
	before	during	after	before	during	after	before	during	after
SC-MSGARCH	6.9	7.8	−3.8	0.8	9.3	4.5	2.2	16.4	4.6
MC-c-MSGARCH	6.3	5.9	−6.3	6.9	5.9	−6.3	11.7	12.1	−14.9
MC-f-MSGARCH	3.9	4.7	−4.8	1.9	4.6	0.2	−5.9	1.0	−3.3

Notes: $(\hat{s}_t, \hat{z}_t)$=argmax $p(s_t, z_t|\mathcal{F}_{t-1})$, Percentage variance reduction are calculated as the differences of variance of hedged portfolio using OLS estimate and estimated variances of alternative models over variance of hedged portfolio using OLS estimate position multiplied by 100. before, during and after respectively signifies the period before, during and after the 2008/2009 global financial crisis. SC-MSGARCH stands for single chain MS-GARCH; MC-c-MSGARCH stands for constrained Multichain MS-GARCH model; and MC-f-MSGARCH stands for unconstrained Multichain MS-GARCH

From Table 4.6, it appears that Markov-switching models provide more efficient hedge ratios relative to the OLS estimate both before and during the 2008/2009 global financial crisis. The OLS hedge ratio on the other hand seem to perform better than MS-GARCH models after the financial crisis. This observation may be due to the low conditional variance of the markets after the 2008/2009 global financial crisis. Among the MS-GARCH specifications under consideration, the constrained multichain MS-GARCH model provides the most consistent measure of hedging effectiveness across the three different measures of hedge ratios used in the evaluation of the of 4.23. While, the unconstrained multichain MS-GARCH model provides the least hedging effectiveness across the three different measures of hedge ratios used in its evaluation. Furthermore, prior to the financial crisis, the hedging effectiveness obtained using the most probable hedge ratio suggests that the OLS hedge ratio perform better than the unconstrained multichain model. This is in contrast with our observation when the average hedge ratio is applied. This observation suggests that the unconstrained multichain model is flexible enough to detect events that are not apparent when average hedge ratio is applied. Our observation is in line with Sephton [1998] observation who finds that the Regime Switching strategy outperforms both OLS and GARCH strategies in the low variance state, but performs much worser than either strategy in the high variance state. This is an indication that multichain Markov-switching models has the potentials of competing favourable with other time-varying models.

It is worth emphasizing that our measure of hedging effectiveness has been shown to be inadequate in evaluating minimum-variance hedge ratios other than OLS. See Lien [2005] and Lien [2009] for discussion. Based on this, alternative measures of effectiveness may provide better insight into the relative advantages of the multichain regime switching model over other models.

4.4 Conclusion

In this paper, we present and examine the performance of a robust Bayesian MS-GARCH model for determining the time-varying hedge ratios in energy futures markets. We account for parameter uncertainty by considering a Bayesian decision rule. Our results suggest that Markov switching models are capable of improving hedging performance in terms of variance reduction. Subject to further research is the issue arising from model uncertainty and the use Bayesian Model Averaging (BMA). Finally, Ederington hedging effectiveness measure is considered in analysing the out-of-sample performance of the proposed hedging strategies in this paper. However, Lien [2005] has emphasized the inadequacy of the regression R^2 to evaluate minimum-variance hedge ratios other than OLS. Also, the hedging effectiveness calculation does not consider transaction cost. Thus, the superiority of the regime switching hedge strategy limits its use for practical purposes. In this respect, it is our plan to address these issues in further research. First by incorporating transaction cost into our modelling framework and secondly by considering alternative measures of hedging effectiveness such as a utility framework in order to create a balance between variance reduction and incremental transaction.

4.A Proof to Proposition 4.2.1:

$$\begin{aligned} h_t &= \arg\min_{h\in H} Var(RS_t - hRF_t|\mathcal{F}^{\Theta}_{t-1}), \\ &= \arg\min_{h\in H} \left(Var(RS_t|\mathcal{F}^{\Theta}_{t-1}) + h^2 Var(RF_t|\mathcal{F}^{\Theta}_{t-1}) - 2hCov(RS_t, RF_t|\mathcal{F}^{\Theta}_{t-1})\right). \end{aligned} \tag{4.24}$$

where $\mathcal{F}_{t-1}$ denotes the information set available up to time t. Under the normal distributional assumption, neither $Var(RS_t|\mathcal{F}^{\Theta}_{t-1})$, $Var(RF_t|\mathcal{F}^{\Theta}_{t-1})$ nor $Cov(RS_t, RF_t|\mathcal{F}^{\Theta}_{t-1})$ depend on the on RS_t and RF_t. Therefore, our problem reduces to

$$h_t = \arg\min_{h\in H} Var(RS_t|\mathcal{F}^{\Theta}_{t-1}) + h^2 Var(RF_t|\mathcal{F}^{\Theta}_{t-1}) - 2hCov(RS_t, RF_t|\mathcal{F}^{\Theta}_{t-1}). \tag{4.25}$$

Taking the first order derivative with respect to h and equating the resulting expression to zero, we have

$$\begin{aligned} h_t &= \frac{Cov(RS_t, RF_t|\mathcal{F}^{\Theta}_{t-1})}{Var(RF_t|\mathcal{F}^{\Theta}_{t-1})}, \\ &= \frac{Cov(\mu(s_t), RF_t|\mathcal{F}^{\Theta}_{t-1})}{V(RF_t|\mathcal{F}^{\Theta}_{t-1})} + \frac{Cov(\nu(s_t, z_t)RF_t, RF_t|\mathcal{F}^{\Theta}_{t-1})}{V(RF_t|\mathcal{F}^{\Theta}_{t-1})}. \end{aligned} \tag{4.26}$$

where,

$$\begin{aligned} Cov(\nu(s_t, z_t)RF_t, RF_t|\mathcal{F}^{\Theta}_{t-1}) &= E[\nu(s_t, z_t)RF_t^2|\mathcal{F}^{\Theta}_{t-1}] - E[\nu(s_t, z_t)RF_t|\mathcal{F}^{\Theta}_{t-1}]E[RF_t|\mathcal{F}^{\Theta}_{t-1}] \\ &= E[\nu(s_t, z_t)(a(z_t) + \tau_t\zeta_t)^2|\mathcal{F}^{\Theta}_{t-1}] \\ &- E[\nu(s_t, z_t)(a(z_t) + \tau_t\zeta_t)|\mathcal{F}^{\Theta}_{t-1}]E[(a(z_t) + \tau_t\zeta_t)|\mathcal{F}^{\Theta}_{t-1}], \\ &\overset{iid\ \zeta_t}{=} E[\nu(s_t, z_t)(a(z_t)^2 + \tau_t^2)|\mathcal{F}^{\Theta}_{t-1}] - E[\nu(s_t, z_t)a(z_t)|\mathcal{F}^{\Theta}_{t-1}]E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}], \\ &= E[\nu(s_t, z_t)(a(z_t)^2 + \tau_t^2 - a(z_t)E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])|\mathcal{F}^{\Theta}_{t-1}]. \end{aligned} \tag{4.27}$$

Then by law of iterated expectation, we have

$$
\begin{aligned}
&Cov(\nu(s_t, z_t)RF_t, RF_t|\mathcal{F}^{\Theta}_{t-1})\\
&= E\left(\sum_{i,j=1}^{M} \nu_{ij}(a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])p(s_t = i, z_t = j|(s,z)_{1:t-1}, \mathcal{F}_{t-1}, \theta)|\mathcal{F}^{\Theta}_{t-1}\right)\\
&= E\left(\sum_{i,j=1}^{M} \nu_{ij}(a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}|\mathcal{F}^{\Theta}_{t-1}\right)\\
&= \sum_{(s,z)_{1:t-1}} \left(\sum_{i,j=1}^{M} \nu_{ij}(a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}\right) p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta)\\
&= \sum_{i,j=1}^{M} \nu_{ij} \left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta)\right),
\end{aligned} \tag{4.28}
$$

where $(s,z)_{s:t} = \{(s_r, z_r)\}_{r=s:t}$ and $\pi_{ij,..} = p(s_t = i, z_t = j|s_{t-1}, z_{t-1}, \theta)$.

Analogously,

$$
V(RF_t|\mathcal{F}^{\Theta}_{t-1}) = \sum_{i,j=1}^{M} \left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta)\right), \tag{4.29}
$$

$$
E(a(z_t)|\mathcal{F}^{\Theta}_{t-1}) = \sum_{(s,z)_{1:t-1}} \sum_{i,j=1}^{M} a_j \pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta),
$$

$$
\begin{aligned}
&Cov(\mu(s_t), RF_t|\mathcal{F}^{\Theta}_{t-1}) = Cov(\mu(s_t), a(z_t)|\mathcal{F}^{\Theta}_{t-1}) =\\
&\qquad = \sum_{(s,z)_{1:t-1}} \sum_{i,j=1}^{M} \left(\mu_i a_j - \mu_i E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}]\right) \pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta),
\end{aligned}
$$

$\tau_t^2(j) = \kappa_j + \omega_j \xi_{t-1}^2 + \phi_j \tau_{t-1}^2$ for $j = 1, \ldots, M$ and t=1,..., T. The result follows immediately by substituting these quantities into (4.26) and letting

$$
w_{ij} = \frac{\left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta)\right)}{\sum_{i,j=1}^{M} \left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta)\right)}.
$$

4.B Proof to Proposition 4.2.5:

$$h_t = \underset{h \in H}{\arg\min} E[(Var(RS_t - hRF_t|\mathcal{F}^{\Theta}_{t-1}))|\mathcal{F}_{t-1}],$$
$$= \underset{h \in H}{\arg\min} E(\big(Var(RS_t|\mathcal{F}^{\Theta}_{t-1}) + h^2 Var(RF_t|\mathcal{F}^{\Theta}_{t-1}) - 2hCov(RS_t, RF_t|\mathcal{F}^{\Theta}_{t-1})\big)\, ||\mathcal{F}_{t-1}). \tag{4.30}$$

where $\mathcal{F}_{t-1}$ denotes the information set available up to time $t-1$. Under the normal distributional assumption, neither $Var(RS_t|\mathcal{F}^{\Theta}_{t-1})$, $Var(RF_t|\mathcal{F}^{\Theta}_{t-1})$ nor $Cov(RS_t, RF_t|\mathcal{F}^{\Theta}_{t-1})$ depend on the on RS_t and RF_t. Therefore, our problem reduces to

$$h_t = \underset{h \in H}{\arg\min} E[(Var(RS_t|\mathcal{F}^{\Theta}_{t-1}) + h^2 Var(RF_t|\mathcal{F}^{\Theta}_{t-1}) - 2h_t Cov(RS_t, RF_t|\mathcal{F}^{\Theta}_{t-1}))|\mathcal{F}_{t-1}]. \tag{4.31}$$

Under some regularity conditions, the first order condition with respect to h_t is given by

$$\begin{aligned} h_t &= \frac{E(Cov(RS_t, RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}{E(Var(RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}, \\ &= \frac{E(Cov(\mu(s_t), RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}{E(V(RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})} + \frac{E(Cov(\nu(s_t, z_t)RF_t, RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}{E(V(RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}. \end{aligned} \tag{4.32}$$

where,

$$\begin{aligned} &E(Cov(\nu(s_t, z_t)RF_t, RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1}) \\ &= \int_{\Theta} \left(\sum_{i,j=1}^{M} \nu_{ij} \left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}]) \pi_{ij,..} p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta) \right) \right) p(\theta|\mathbf{y}_{1:t-1}) d\theta \\ &= \sum_{i,j=1}^{M} \left(\int_{\Theta} \nu_{ij} \left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}]) \pi_{ij,..} p((s,z)_{1:t-1}|\mathcal{F}_{t-1}, \theta) \right) p(\theta|\mathbf{y}_{1:t-1}) d\theta \right) \\ &= \sum_{i,j=1}^{M} \left(\int_{\Theta} \nu_{ij} \left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}]) \pi_{ij,..} p((s,z)_{1:t-1}, \theta, |\mathcal{F}_{t-1}, \theta) \right) d\theta \right), \end{aligned} \tag{4.33}$$

and

$$
\begin{aligned}
&E(V(RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})\\
&=\int_{\Theta}\left(\sum_{i,j=1}^{M}\left(\sum_{(s,z)_{1:t-1}}(a_j^2+\tau_t^2(j)-a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1},\theta)\right)\right)p(\theta|\mathbf{y}_{1:t-1})d\theta\\
&=\sum_{i,j=1}^{M}\left(\int_{\Theta}\left(\sum_{(s,z)_{1:t-1}}(a_j^2+\tau_t^2(j)-a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1},\theta)\right)p(\theta|\mathbf{y}_{1:t-1})d\theta\right)\\
&=\sum_{i,j=1}^{M}\left(\int_{\Theta}\left(\sum_{(s,z)_{1:t-1}}(a_j^2+\tau_t^2(j)-a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1},\theta|\mathcal{F}_{t-1},\theta)\right)d\theta\right).
\end{aligned}
\tag{4.34}
$$

$\tau_t^2(j) = \kappa_j + \omega_j\xi_{t-1}^2 + \phi_j\tau_{t-1}^2$ for $j = 1,\dots,M$ and t=1,..., T. The result follows immediately by substituting these quantities into (4.32) with

$$
\begin{aligned}
&w_{ij}(\theta|\mathbf{y}_{1:t-1})\\
&=\frac{\left(\sum_{(s,z)_{1:t-1}}(a_j^2+\tau_t^2(j)-a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1},\theta|\mathcal{F}_{t-1},\theta)\right)}{\sum_{i,j=1}^{M}\left(\int_{\Theta}\left(\sum_{(s,z)_{1:t-1}}(a_j^2+\tau_t^2(j)-a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1},\theta|\mathcal{F}_{t-1},\theta)\right)d\theta\right)}.
\end{aligned}
$$

4.C Constructing proposal distribution for $\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau$

Sample $(\theta_u^{RS})^{(g)}, (\theta_a^{RF})^{(g)}, \theta_\sigma^{(g)}, \theta_\tau^{(g)}$ from $f(\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau|s_{1:t}^{(g)}, \pi^{(g)}, y_{1:t})$. Given a prior density $f(\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau)$, the posterior density of $(\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau)$ can be expressed as follows

$$
\begin{aligned}
f(\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau|s_{1:t}^{(r)}, \pi, y_{1:t}) \propto{}& f(\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau)\\
&\times\prod_{i=1}^{t}\frac{1}{\sigma_i}\exp\left(-\frac{(RS_i-\mu(s_i)-\nu(s_i)RF_i)^2}{2\sigma_i^2}\right)\\
&\times\prod_{i=1}^{i}\frac{1}{\tau_i}\exp\left(\frac{-(RF_i-a(s_i))^2}{2\tau_i^2}\right)
\end{aligned}
\tag{4.35}
$$

where,

$$\sigma_i^2 = \gamma(s_i) + \alpha(s_i)(RS_{i-1} - \mu(s_{i-1}) - \nu(s_{i-1})RF_{i-1})^2 + \beta(s_i)\sigma_{i-1}^2.$$

and

$$\tau_i^2 = \kappa(s_i) + \omega(s_i)(RF_{i-1} - a(s_{i-1}))^2 + \phi(s_i)\tau_{i-1}^2.$$

In order to generate $\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau$ from the joint distribution we first separate the parameters of RS_t from RF_t and apply further blocking on this subgroups of the Gibbs sampler i.e. We split the regime-dependent parameters of both RS_t and RF_t into two subvectors, the parameter of the observation equation θ_u^{RS} (θ_a^{RF}) and the parameters of the volatility process θ_σ (θ_τ). For each subvector we implement a Metropolis-Hastings (MH) step that samples from normal distribution in the case of θ_u^{RS} (θ_a^{RF}) and truncated normal distribution in the case of θ_σ (θ_τ). The distributions is adapted during the burnin period.

As regards the parameters of the conditional expectation of the θ_u^{RS}, we derive the mean and variance of the proposal distribution by considering an approximation of the full conditional distribution of θ_u^{RS},

$$f(\theta_u^{RS}|s_{1:t}^{(g)}, \gamma^{(g-1)}, \beta^{(g-1)}, \alpha^{(g-1)}, RS_{1:t}, RF_{1:t}) \propto \prod_{i=1}^{t} \mathcal{N}(RS_i; \mu(s_i) + \nu(s_i)RF_i, \sigma_i^2).$$

Given an approximation σ_t^{*2} of σ_t^2, it can easily be shown, by completing the square method, that the full conditional distribution of θ_u^{RS} can be approximated by a normal distribution. Let

$$\nabla_{ut} = \begin{pmatrix} 1 & 0 & \cdots & 0 & RF_t & 0 & \cdots & 0 \\ 0 & 1 & 0 & \vdots & 0 & RF_t & 0 & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & 1 & 0 & \cdots & 0 & RF_t \end{pmatrix}',$$

$$\mathbf{V_u} = \begin{pmatrix} \sigma_1^{*2} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_t^{*2} \end{pmatrix},$$

and define a $t \times 2M$ matrix ∇_u whose $i-$th row corresponds to $\nabla_{ui}\xi_i$ and $\xi_i = (\mathbb{I}_{s_i=1}, \dots, \mathbb{I}_{s_i=M})'$ then

$$\begin{aligned} & f(\theta_u^{RS}|s_{1:t}^{(g)}, \gamma^{(g-1)}, \beta^{(g-1)}, \alpha^{(g-1)}, RS_{1:t}, RF_{1:t}) \\ & \approx \frac{1}{|\mathbf{V}|^{\frac{1}{2}}} \exp\left(-\frac{(\mathbf{RS'_{1:t}} - \nabla_\mathbf{u}\theta_\mathbf{u}^{\mathbf{RS'}})'\mathbf{V_u^{-1}}(\mathbf{RS'_{1:t}} - \nabla_\mathbf{u}\theta_\mathbf{u}^{\mathbf{RS'}})}{2} \right) \\ & = \mathcal{N}_{2M}(m_u, \Sigma_u), \end{aligned}$$

where,

$$\Sigma_u = (\nabla_u' \mathbf{V_u}^{-1} \nabla_u)^{-1}$$
$$m_u = \Sigma_u \nabla_u' \mathbf{V_u}^{-1} RS_{1:t}'.$$

$$\sigma_i^{*2} = \gamma^{(g-1)}(s_i^{(g)}) + \alpha^{(g-1)}(s_i^{(g)})(RS_{i-1} - \mu^{(g-1)}(s_{i-1}^{(g)}) - \nu^{(g-1)}(s_{i-1}^{(g)})RF_{i-1})^2 + \beta^{(g-1)}(s_i^{(g)})\sigma_{i-1}^2.$$

As regards the parameters of the volatility process the full conditional is

$$f(\theta_u^{RS}|s_{1:t}^{(g)}, \gamma^{(g-1)}, \beta^{(g-1)}, \alpha^{(g-1)}, RS_{1:t}, RF_{1:t}) \propto \prod_{i=1}^{t} \mathcal{N}(RS_i; \mu(s_i) + \nu(s_i)RF_i, \sigma_i^2).$$

We now follow the ARMA approximation of the MS-GARCH process i.e.

$$\begin{aligned} \sigma_t^2 &= \gamma(s_t) + \alpha(s_t)\epsilon_{t-1}^2 + \beta(s_t)\sigma_{t-1}^2 \\ \epsilon_t^2 &= \gamma(s_t) + (\alpha(s_t) + \beta(s_t))\epsilon_{t-1}^2 - \beta(s_t)(\epsilon_{t-1}^2 - \sigma_{t-1}^2) + (\epsilon_t^2 - \sigma_t^2). \end{aligned}$$

Let

$$w_t = \epsilon_t^2 - \sigma_t^2 = \left(\frac{\epsilon_t^2}{\sigma_t^2} - 1\right)\sigma_t^2 = (\chi^2(1) - 1)\sigma_t^2$$

with

$$E_{t-1}[w_t] = 0; \quad \text{and} \quad Var_{t-1}[w_t] = 2\sigma_t^4.$$

Subject to the above and following Nakatsuma [1998] suggestion, we assume that $w_t \approx w_t^* \sim \mathcal{N}(0, 2\sigma_t^4)$. Then we have an "auxiliary"ARMA model for the squared error ϵ_t^2.

$$\begin{aligned} &\epsilon_t^2 = \gamma(s_t) + (\alpha(s_t) + \beta(s_t))\epsilon_{t-1}^2 - \beta(s_t) w_{t-1}^* + w_t^*, \quad w_t^* \sim \mathcal{N}(0, 2\sigma_t^4) \\ \text{i.e.} \quad & w_t^* = \epsilon_t^2 - \gamma(s_t) - \alpha(s_t)\epsilon_{t-1}^2 - \beta(s_t)(\epsilon_{t-1}^2 - w_{t-1}^*). \end{aligned} \tag{4.36}$$

Following Ardia [2008] we further express w_t^* as a linear function of $(3M \times 1)$ vector of $\theta_\sigma = (\gamma_1, \ldots, \gamma_M, \alpha_1, \ldots, \alpha_M, \beta_1, \ldots, \beta_M)'$. To do this, we approximate the function w_t^* by first order Taylor's expansion about $\theta_\sigma^{(r-1)} = (\gamma_1^{(r-1)}, \ldots, \gamma_M^{(r-1)}, \alpha_1^{(r-1)}, \ldots, \alpha_M^{(r-1)}, \beta_1^{(r-1)}, \ldots, \beta_M^{(r-1)})'$.

$$w_t^* \approx w_t^{**} = w_t^*(\theta_{-\pi}^{(r-1)}) - (\theta_\sigma - \theta_\sigma^{(r-1)})' \nabla_t \xi_t,$$

where

$$\nabla_t = - \begin{pmatrix} \frac{\partial w_t^*}{\partial \gamma_1} & 0 & \cdots & 0 & \frac{\partial w_t^*}{\partial \alpha_1} & 0 & \cdots & 0 & \frac{\partial w_t^*}{\partial \beta_1} & 0 & \cdots & 0 \\ 0 & \frac{\partial w_t^*}{\partial \gamma_2} & 0 & \vdots & 0 & \frac{\partial w_t^*}{\partial \alpha_2} & 0 & \vdots & 0 & \frac{\partial w_t^*}{\partial \beta_2} & 0 & \vdots \\ \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 & \vdots & 0 & \ddots & 0 \\ 0 & \cdots & 0 & \frac{\partial w_t^*}{\partial \gamma_M} & 0 & \cdots & 0 & \frac{\partial w_t^*}{\partial \alpha_M} & 0 & \cdots & 0 & \frac{\partial w_t^*}{\partial \beta_M} \end{pmatrix}'$$

and

$$\begin{aligned} \frac{\partial w_t^*}{\partial \gamma_k} &= -\xi_{tk} + (\xi_t' \beta) \frac{\partial w_{t-1}^*}{\partial \gamma_k} \\ \frac{\partial w_t^*}{\partial \alpha_k} &= -\xi_{tk}\epsilon_{t-1}^2 + (\xi_t' \beta) \frac{\partial w_{t-1}^*}{\partial \alpha_k} \\ \frac{\partial w_t^*}{\partial \beta_k} &= -\xi_{tk}(\epsilon_{t-1}^2 - w_{t-1}^*) + (\xi_t' \beta) \frac{\partial w_{t-1}^*}{\partial \beta_k} \end{aligned}$$

for $k = 1, \ldots, M$, evaluated at $\theta_\sigma^{(r-1)}$.

Upon defining $r_t^* = w_t^*(\theta_{-\pi}^{(r-1)}) + \theta_\sigma^{'(r-1)} \nabla_t \xi_t$, it turns out that $w_t^{**} = r_t^* - \theta_\sigma' \nabla_t \xi_t$. Furthermore, by defining the $T \times 1$ vectors $\mathbf{w} = (w_1^{**}, \ldots, w_T^{**})'$, $\mathbf{r}^* = (r_1^*, \ldots, r_T^*)'$, a

$T \times 3M$ matrix ∇ whose $t-$th row corresponds to $\xi_t' \nabla_t'$ as well as a $T \times T$ matrix

$$\mathbf{V} = 2 \begin{pmatrix} \sigma_1^{**4} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma_T^{**4} \end{pmatrix},$$

with $\sigma_t^{**2} = (\xi_t^{(r)'} \gamma^{(r-1)}) + (\xi_t^{(r)'} \alpha^{(r-1)})(y_{t-1} - \xi_{t-1}^{(r)}{}' \mu^{(r)})^2 + (\xi_t^{(r)'} \beta^{(r-1)}) \sigma_{t-1}^{**2}$, we end up with $\mathbf{w} = \mathbf{v} - \theta_\sigma' \nabla$. Using this linear approximation, we can approximate the full conditional probability of the volatility parameters as

$$\begin{aligned} & f(\theta_\sigma | \xi_{1:T}^{(r)}, \mu^{(r)}, y_{1:T}) \propto \\ & \propto \frac{1}{|\mathbf{V}|^{\frac{1}{2}}} \exp\left(-\frac{\mathbf{w}'\mathbf{V}^{-1}\mathbf{w}}{2} \right) \mathbb{I}_{\{\gamma_1 > 0 \ldots, \gamma_M > 0, 0 < \alpha_1 < 1, \ldots, 0 < \alpha_M < 1, 0 < \beta_1 < 1 \ldots, 0 < \beta_M < 1\}} \\ & \propto \mathcal{N}_{3M}(m_\sigma, \Sigma_\sigma) \mathbb{I}_{\{\gamma_1 > 0 \ldots, \gamma_M > 0, 0 < \alpha_1 < 1, \ldots, 0 < \alpha_M < 1, 0 < \beta_1 < 1 \ldots, 0 < \beta_M < 1\}}, \end{aligned} \tag{4.37}$$

where

$$\begin{aligned} \Sigma_\sigma &= (\nabla' \mathbf{V}^{-1} \nabla)^{-1} \\ m_\sigma &= \Sigma \nabla' \mathbf{V}^{-1} \mathbf{r}^*. \end{aligned} \tag{4.38}$$

In a similar fashion we construct the proposal distribution for the parameters of RF_t

Bibliography

A. H. Alizadeh and N. K. Nomikos. A Markov regime switching approach for hedging stock indices. *Journal of Futures Markets*, 24:649 – 674, 2004.

A. H. Alizadeh, N. K. Nomikos, and P. K. Pouliasis. A Markov regime switching approach for hedging energy commodities. *Journal of Banking & Finance*, 32:1970 – 1983, 2008.

Billy Amzal, Frédéric Y Bois, Eric Parent, and Christian P Robert. Bayesian-optimal design via interacting particle systems. *Journal of the American Statistical Association*, 101(474):773–785, 2006.

D. Ardia. *Financial Risk Management with Bayesian Estimation of GARCH Models: Theory and Applications, volume 612 of Lecture Notes in Economics and Mathematical Systems.* Springer-Verlag, Berlin, Germany, 2008.

V.S. Bawa, S.J. Brown, and R.W. Klein. *Estimation Risk and Optimal Portfolio Choice*. Studies in Bayesian econometrics. North-Holland Publishing Company, 1979.

M. Billio, R. Casarin, and A. Osuntuyi. Efficient Gibbs sampling for Markov switching GARCH models. 2012.

C. Y. Chang, J. Y. Lai, and I. Y. Chuang. Futures hedging effectiveness under the segmentation of bear/bull energy markets. *Energy Economics*, 32:442 – 449, 2010.

S. Chen, C. Lee, and K. Shrestha. Futures hedge ratios: a review. *The Quarterly Review of Economics and Finance*, 43:433 – 465, 2003.

Morris H DeGroot. *Optimal statistical decisions*, volume 82. Wiley-Interscience, 2005.

A. Doucet, N. de Freitas, and N. J. Gordon. *Sequential Monte Carlo methods in practice*, volume 1. Springer New York, 2001.

L.H. Ederington. The hedging performance of the new futures markets. *Journal of Finance*, 34(1):157 – 170, 1979.

S. F. Gray. Modeling the conditional distribution of interest rates as a regime-switching process. *Journal of Financial Economics*, 42:27–62, 1996.

M. S. Haigh and M. T. Holt. Crack spread hedging: accounting for time-varying volatility spillovers in the energy futures markets. *Journal of Applied Econometrics*, 17:269 – 289, 2002.

L. L. Johnson. The theory of hedging and speculation in commodity futures. *Review of Economic Studies*, 27(3):139 – 151, 1960.

R. Kan and G. Zhou. Optimal portfolio choice with parameter uncertainty. *Journal of Financial and Quantitative Analysis*, 42(3):621, 2007.

H. Lee and J. K. Yoder. A bivariate Markov regime switching GARCH approach to estimate time varying minimum variance hedge ratio. *Applied Economics*, 39(10): 1253 – 1265, 2007a.

H. Lee and J. K. Yoder. Optimal hedging with a regime switching time varying correlation GARCH model. *Journal of Futures Markets*, 27(5):495 – 516, 2007b.

H. T. Lee, J. K. Yoder, R. C. Mittelhammer, and J. J. McCluskey. A random coefficient autoregressive Markov regime switching model for dynamic futures hedging. *The Journal of Futures Markets*, 26(2):103 – 129, 2006.

S. H. Lence and D. J. Hayes. The empirical minimum-variance hedge. *American Journal of Agricultural Economics*, 76(1):94 – 104, 1994a.

S. H. Lence and D. J. Hayes. Parameter-based decision marketing under estimation risk: An application to futures trading. *Journal of Finance*, 49(1):345 – 357, 1994b.

D Lien. A note on the superiority of the ols hedge ratio. *Journal of Futures Markets*, 25:11211126, 2005.

D Lien. A note on the hedging effectiveness of GARCH models. *International Reviews of Economics and Finance*, 18:110 – 112, 2009.

P. Müller. Simulation-based optimal design. *Bayesian statistics*, 6:459–474, 1999.

P. Müller, B. Sansó, and M. De Iorio. Optimal bayesian design by inhomogeneous markov chain simulation. *Journal of the American Statistical Association*, 99(467): 788–798, 2004.

R. Myers. Estimating time-varying optimal hedge ratios on futures markets. *Journal of Futures Markets*, 11:39 – 53, 1991.

T. Nakatsuma. A Markov-chain sampling algorithm for GARCH models. *Studies in Nonlinear Dynamics and Econometrics*, 3(2):107–117, 1998.

H. Raiffa and R. Schlaifer. Applied statistical decision theory (harvard business school publications). 1961.

P. Sephton. Garch and markov hedging at the winnipeg commodity exchange. *Canadian Journal of Agricultural Economics*, 46:117126., 1998.

H.-J. Sheu and H.-T. Lee. Optimal futures hedging under multichain markov regime switching. *Journal of Futures Markets*, 2012.

Yun Yang and David B Dunson. Sequential markov chain monte carlo. *arXiv preprint arXiv:1308.3861*, 2013.

Chapter 5

Bayesian Approach to Forecast Rationality Tests of A.J. Patton and A. Timmermann

with Lennart Hoogerheide, and Herman K. van Dijk[1]

Abstract We propose Bayesian inference as an alternative procedure for testing the monotonicity properties of second moment bounds across several horizons as presented in Patton and Timmermann [2012]. This approach avoids the use of computationally expensive test statistics and asymptotic distributions under the null hypothesis and consequently producing more reliable results for moderate and small sample data. We are also able to account for parameter uncertainty using our proposed Bayesian inference technique.

Keywords : Markov Switching, Hedge ratio, Energy futures, GARCH

[1]This paper was developed during my research visit to Tinbergen Institute, Vrije Universiteit Amsterdam, The Netherlands.

5.1 Introduction

Forecasts are used by policymakers and other practitioners for making informed decisions in their businesses. For example, Government's decision on taxes, revenues, money and credit supply, foreign trade and balances, employment among others rests upon actual or implied forecasts of economic conditions. In finance, on the other hand, forecast plays important role in such studies as portfolio theory, asset pricing and corporate among others. Forecasts on various economic and financial variables are published by bodies like the Philadelphia Federal Reserve (Survey of Professional Forecasters) and the IMF (World Economic Outlook). These forecasts are usually generated using econometric models and/or subjective judgment; see Ross [1955] for a survey on economic forecasting techniques. A stream of research on forecasting takes the view of a forecast producer and focuses on the issues of constructing optimal forecast subject to a loss function. See Elliot and Timmermann [2004], Geweke and Whiteman [2006], Zellner [1986], Hashem Pesaran et al. [2013] for illustration. The other stream of research takes on the forecast evaluator's perspective and deals with issues surrounding tests of forecast rationality/optimality. Forecasts on a single economic or financial variable are usually published by several independent forecasters and, generally, the underlying forecasting processes used for generating these forecasts are not made known. Hence, judging whether these available forecasts are good or not is vital in the decision making process of end users. In this paper, we focus on forecast rationality test.

As proposed by Muth [1961] and further developed by Nelson [1975] and Sargent [1972], we say forecasts are rational if they fully incorporate all the information available at the time they are made. Two main properties (i.e. efficiency and consistency) are placed on forecasts by this definition. The first property, efficiency, requires that in generating one-period ahead forecast, forecasters should utilize all the information in the realized rates of the variable being forecast. In essence, one-period ahead forecast and realization share the same autoregressive pattern. The other property, consistency, demands that all the available information should be consistently applied in producing multi-horizon forecast. Thus, forecasts are consistent if multi-horizon forecasts are obtained recursively, with the yet to be observed realizations of the series replaced by rational forecast.

The efficiency and consistency requirement of rational forecast, so described here, must hold under quite general conditions. More particularly, these properties must be satisfied (see Modigliani and Shiller [1973]) if the forecast are optimal in the least square (quadratic loss function) sense. Based on this, tests of forecast rationality are usually constructed by first defining an optimal forecast subject to a loss function and then use the properties of this optimal forecast to evaluate it. Diebold and Lopez [1996] and Granger and Newbold [1986] show that under covariance stationarity and Mean Squared Error (MSE) loss function, rational forecast is unbiased, one-period forecast errors are serially uncorrelated, $h-$step-ahead forecast error exhibit zero serial correlation beyond lag $(h-1)$ and the unconditional variance of the forecast error is a weakly increasing function of the forecast horizon. Two of these properties (i.e. forecast errors have zero mean and are uncorrelated with any information available at the time that the forecast is made) are routinely checked when testing for forecast rationality. Given a set of forecasts, $y_{t|t-h}$, for a variable y_t, a commonly used test procedure for checking for unbiasedness follows the proposal by Mincer and Zarnowitz [1969] i.e. a simple regression $y_t = a + \beta y_{t|t-h} + e_t$ is carried out. If the estimated value of a and β are respectively not significantly different from 0 and 1, then the forecasts are unbiased. The efficiency test, on the other hand, requires determining if the forecast errors $(y_t - y_{t|t-h})$ are systematically related to some other variable that was known to the forecaster when the forecasts were made. Suppose this other variable is denoted Z_t. Then the efficiency test requires running a regression of the form $(y_t - y_{t|t-h}) = g + dZ_t + u_t$. If the estimated value of d is significantly different from 0, the forecasts are inefficient. Unfortunately, some of the properties of optimal forecast under covariance stationarity and symmetric forecast error loss function assumption do not provide useful guideline for empirical tests because they do not generally hold under other loss function (Patton and Timmermann [2007b]) or the data generating process of the variable of interest does not satisfy the covariance stationarity assumption (Giacomini and Rossi [2010]). In Giacomini and Rossi [2010] and Rossi and Sekhposyan [2011] the assumption of covariance stationarity is dropped and methods for performing inference on forecast comparisons when the forecasting ability may be affected by instabilities is developed. Granger and Newbold [1986] on the other hand argue that the assumption of symmetric loss function does not provide a good representation of a forecaster's objective. Consequently, forecast rationality tests in Economics and Finance are therefore derived by either assuming a specific asymmetric loss function (see, for example, Zellner [1986], Patton and Timmermann [2007b], Aretz et al. [2011]) or imposing testable restrictions on the data generating process of the variable that is being forecast (see Patton and Timmermann [2007a], Elliot et al. [2005] for illustration). Test of forecast rationality is extended to the multivariate setting in Komunjer and Owyang [2012].

In contrast to the one-step ahead forecast upon which the tests of forecast rationality discussed above are designed, empirical records on forecasts are usually produced on two or more horizons at the same time. This observation suggests that the failure to include this information when assessing the quality of forecasts could lead to wrong conclusion that the forecasts are rational. To this end, Capistran [2007] propose tests of forecast rationality based on the property that the unconditional variance of the forecast error under MSE loss function is a non decreasing function of the forecast horizon. More recently, Patton and Timmermann [2012] propose several forecast rationality tests arising from such implications of forecast rationality as; the mean squared forecast error (MSFE) should be non-decreasing with the forecast horizon

(Diebold, 2001, and Patton and Timmermann [2007b]) and that the mean squared forecast should be decreasing with the horizon. They also propose new regression based tests of forecast rationality called "optimal revision regression". This regression model uses the complete set of multi-horizon forecasts in a univariate regression. One other novelty of their paper is that the proposed forecast rationality tests can be implemented without observing the target variable. Patton and Timmermann [2012] further show that both their monotonicity results and "optimal revision regression"test hold when the target variable is replaced with the short horizon forecast. Following Patton and Timmermann [2012], Rossi [2012] propose a robust forecast rationality test that can account for the presence of instability in the DGP of the target variable.

Our suggestion to the literature on tests of forecast rationality are manifolds. First, we propose an easy to compute Bayesian based procedure to illustrate how to implement the joint tests of inequality constraint described in Patton and Timmermann [2012]. In Patton and Timmermann [2012], the authors suggest testing the monotonicity properties via tests of inequality constraints using the methods of Gourieroux et al. [1982], Wolak [1987] and Wolak [1989], and also the bootstrap methods of White [2003] and Hansen [2005]. All these methods are quite challenging to implement and they rely on the asymptotic properties of the test statistics which may not produce reliable results in small sample situation. The second contribution involves the use of our Bayesian inference procedure to account for parameter uncertainties when evaluating the joint tests of the inequality constraints described in Patton and Timmermann [2012]. In Patton and Timmermann [2012] the inequality tests procedure are implemented under the assumption that the parameter value used in the test procedure are known by the forecasters. This approach totally ignores parameter estimation risk. Hoogerheide et al. [2012] provides an interesting application of the "optimal revision regression"to Value-at-Risk forecasting. This application played on the fact that observation on the target variable are not required, a desirable feature in volatility or Value-at-risk applications, where the target variable is latent. They further show, through a simulation study, that when parameter estimation error are ignored the test results are oversized. A simulation based method for obtaining critical values that take into account parameter uncertainty was then propose by the authors. The third contribution involves an application of our test procedure to the Livingston Survey's stock market forecast. How well forecasters forecast the stock market index has been examined in a number of articles. Early studies on this dataset by Lakonishok [1980] and Pearce [1984] found that the forecasts were biased and inefficient. Contrary to this result, Dokko and Edelstein [1989] reached a conclusion that there is no evidence against the rationality of this forecasts. He further argued that the results obtained by both earlier studies were a reflection of wrong computation of expected returns on the stock market index since survey participants are not required to provide forecast for the current June or December value of the stock market index. In previous studies, a common assumption made, for example in the June survey, is that forecasts were made on a particular day. This assumptions is faulted as different participants send in their forecasts on different days and the day-to-day movements of the stock market may be large thus influencing different choices of base index upon which survey participant make their forecast. These findings suggest that the results of all these statistical tests are very sensitive to researchers assumptions. A more recent study on the forecast rationality of stock market index is carried out by Aretz et al. [2011]. In their paper, they combined the approaches of Elliot et al. [2005] and Patton and Timmermann [2007a] with a block bootstrap to show that allowing for variation in the asymmetric loss functions across forecasters, forecasts rationality may no longer be often rejected. Other studies that have been conducted on the Livingston survey data on the forecast of S&P 500 stock market index include stock returns predictability (Söderlind [2010]), herding bahaviour of forecasters (Pierdzioch and Rülke [2012]) and test of behavioural finance theories and asset pricing models (De Bondt [1991]).

The rest of the paper is as follows. In Section 2 we describe the methodology for evaluating hypothesis of inequality constraints. Section 3 presents results from simulation exercise. Section 4 presents our results on empirical exercise and section 5 concludes.

5.2 Evaluating Hypothesis of Inequality Constraints

Following the highlighted implications, i.e. monotonicity and covariance bounds properties, of forecast rationality under squared error loss function in Patton and Timmermann [2012], the authors suggest and illustrate test of these implications through inequality constraint methods of Gourieroux et al. [1982], Wolak [1987, 1989] and the bootstrap methods of White [2003] and Hansen [2005]. These tests approaches rely on asymptotic distribution and asymptotic approximation of the test statistics which may be complicated and challenging to implement. Based on these inherent challenges associated with classical test procedures to inequality constraints, we take up a Bayesian approach to the problem of inequality constraints. This approach does not require the use of complicated test statistics and works well in small or medium sample situation. Bayesian inference on econometric models with inequality constraints on the coefficients have been well studied in the literature. Geweke [1995], Davis [1978], Chamberlain and Leamer [1976], among others, approached the problem of inequality constraint from a Bayesian perspective by analysing linear regression models subject to linear inequality constraints on the coefficients. These literatures defer in the number of inequality or mixed equality and inequality constraints involved. Geweke [1986] extends the study to linear models subject to nonlinear inequality constraints on the coefficients and implements a Monte Carlo integration technique for evaluating the expected value of the function of interest. In this paper, we approach the problem of evaluating hypothesis of inequality constraints by using the "crude frequency simulation" method proposed by Geweke [1986].

Suppose $Y_t = (y_t, y_{t|t-1}, \ldots, y_{t|t-H})'$ is an observation from a random vector in $\Re^{(H+1)}$ with probability density function $f(Y_t, \theta)$, where θ represents the vector of unknown parameters and the function $f(Y_t, \theta)$ is a continuous function of θ for all Y_t with $t = 1, 2, \ldots, T$. The vector of observation is a collection of the target variable at time t and its forecast made at h $(h = 1, 2, \ldots, H)$ horizons before. Let the inequality constraints be expressed as

$$\mathbf{l} \leq \mathbf{g}(\theta) \leq \mathbf{u}, \tag{5.1}$$

where $\mathbf{g}(\theta) = (g_1(\theta), \ldots, g_p(\theta))'$, is a set of real valued function on θ, while $\mathbf{l} = (l_1, \ldots, l_p)'$ and $\mathbf{u} = (l_1, \ldots, l_p)'$ respectively represents the lower and upper bounds on $\mathbf{g}(\theta)$. Elements of $\mathbf{l}$ and $\mathbf{u}$ are allowed to be equal to $\pm\infty$.

Under the above specification, we combine a non-informative prior, $\pi(\theta)$, on θ with the likelihood function $(\mathcal{L}(\theta|Y_1', \ldots, Y_T')$, to obtain the following posterior probability when the parameters are unconstrained

$$f_u(\theta|Y_1', \ldots, Y_T') = \frac{\mathcal{L}(\theta|Y_1', \ldots, Y_T')\pi(\theta)}{\int \mathcal{L}(\theta|Y_1', \ldots, Y_T')\pi(\theta)d\theta}. \tag{5.2}$$

In the constrained case, we define the prior probability to be equal to the product of a non-informative distribution, $\pi(\theta)$, and an indicator function $\mathbb{I}_{\theta \in Q}$ $(Q = \{\theta : \mathbf{l} \leq \mathbf{g}(\theta) \leq \mathbf{u}\})$ representing the inequality constraints. The indicator function $\mathbb{I}_{\theta \in Q} = 1$ if $\theta \in Q$ and $\mathbb{I}_{\theta \in Q} = 0$ if $\theta \notin Q$. The posterior distribution is then defined by replacing the prior probability in Equation (5.2) with the modified prior. i.e.

$$f_c(\theta|Y_1', \ldots, Y_T') = \frac{\mathcal{L}(\theta|Y_1', \ldots, Y_T')\pi(\theta)\mathbb{I}_{\theta \in Q}}{\int \mathcal{L}(\theta|Y_1', \ldots, Y_T')\pi(\theta)\mathbb{I}_{\theta \in Q}d\theta}. \tag{5.3}$$

Evaluation of the inequality constraints in Equation (5.1) as a formal hypothesis is then carried out by assuming that the econometric model with inequality constraints and that without constraints are two competing hypothesis for which a posterior odds ratio is to be formed. Without loss of generality, we assume that the prior odds ratio is 1:1. To this end, an estimate of the posterior ratio, $p_{1|2}$, is given by

$$\begin{aligned} p_{1|2} &= \frac{\int \mathcal{L}(\theta|Y_1',\ldots,Y_T')\pi(\theta)\mathbb{I}_{\theta\in Q}d\theta}{\int \mathcal{L}(\theta|Y_1',\ldots,Y_T')\pi(\theta)d\theta} \\ &= \int \frac{\mathcal{L}(\theta|Y_1',\ldots,Y_T')\pi(\theta)}{\int \mathcal{L}(\theta|Y_1',\ldots,Y_T')\pi(\theta)d\theta}\mathbb{I}_{\theta\in Q}d\theta \\ &= \int \mathbb{I}_{\theta\in Q} f_u(\theta|Y_1',\ldots,Y_T')d\theta \\ &= E[\mathbb{I}_{\theta\in Q}]. \end{aligned} \tag{5.4}$$

Observe that the posterior odds ratio, $p_{1|2}$, is essentially the expected value of the indicator function ($\mathbb{I}_{\theta\in Q}$) representing the inequality constraints under the posterior distribution of the unconstrained model. Analytical evaluation of this quantity may be a difficult one to handle due to the presence of the indicator function. As a result of this, we approximate the posterior odds ratio by implementing a Monte Carlo integration technique. This is done by generating G random samples of θ from the unconstrained posterior probability and then compute

$$\hat{p}_{1|2} = \sum_{i=1}^{G} \frac{\mathbb{I}_{\theta^i\in Q}}{G}. \tag{5.5}$$

The number of samples needed to achieve a good approximation depends on the complexity of the inequality constraints. The advantage of this procedure lies in its simplicity but it is known to be inefficient and requires many draws, G, if $\hat{p}_{1|2}$ is small. Geweke [1995] propose a more efficient method based on Geweke-Hajivassiliou-Keane (GHK) probability simulator to evaluate the constrained posterior probability within the framework of linear regression models subject to linear inequality constraints on the coefficient. We leave as a subject for further research an extension of this technique to the case of nonlinear inequality constraints.

In the methodology presented above, we assumed a 1:1 prior odds ratio. In general, a substantive result of the posterior odds ratio, $p_{1|2}$, may be obtained by evaluating the limiting ratio formed from the sequence of the prior odds ratio chosen.

Finally, having been able to state the posterior odds and probabilities, we are now left with taking an action: that is, accept H_0 or reject H_0. We address this problem by following the expected utility hypothesis framework described on pages 294 - 297 of Zellner [1996]. Under a two-state-two-action situation, we assume the following loss structure as a representative of the consequences of our action:

Action	State of the world	
	H_0 true	H_1 true
Accept H_0	$L(H_0, \hat{H}_0) = 0$	$L(H_1, \hat{H}_0)$
Accept H_1	$L(H_0, \hat{H}_1)$	$L(H_1, \hat{H}_1) = 0$

where $L()$ represent the loss experienced as a result of our action, $\hat{H}_0$ and $\hat{H}_1$ respectively denote our acceptance of $\hat{H}_0)$ or H_1. In the above loss structure, we assign zero loss to a situation where the action taken corresponds to the state of the world. On the hand, a positive loss $L(H_1, \hat{H}_0)$ is assumed if we accept H_0 when it is actually false.

Based on this loss structure and given the posterior probabilities for the hypothesis H_0 and H_1, we evaluate the consequences of the action taken by computing the expected loss associated with the action accept H_0

$$E(L|\hat{H}_0) = p(H_1|Y)L(H_1, \hat{H}_0),$$

and accept H_1

$$E(L|\hat{H}_1) = p(H_0|Y)L(H_0, \hat{H}_1).$$

Having computed the expected loses, an action is taken by comparing the expected loses as follows:

$$\text{If} \quad E(L|\hat{H}_0) < E(L|\hat{H}_1), \quad \text{accept} \quad H_0, \quad \text{action} \quad \hat{H}_0,$$

$$\text{If} \quad E(L|\hat{H}_1) < E(L|\hat{H}_0), \quad \text{accept} \quad H_1, \quad \text{action} \quad \hat{H}_1.$$

This decision rule can further be expressed in terms of the posterior odds ratio of the hypothesis and the ratio of the loss function i.e. we accept H_0 if

$$\frac{p(Y|H_0)p(H_0)}{p(Y|H_1)p(H_1)} > \frac{L(H_1, \hat{H}_0)}{L(H_0, \hat{H}_1)}.$$

If we further assume a symmetric loss structure i.e.

$$L(H_0, \hat{H}_1) = L(H_1, \hat{H}_0); \qquad L(H_0, \hat{H}_0) = L(H_1, \hat{H}_1) = 0$$

then, we accept H_0 if the posterior odds ratio is greater than 1 and H_1 otherwise.

In the following section, we present a simulation exercise involving the application of our propose Bayesian inference technique for inequality constraints to the problem of forecast rationality based on monotonicity and covariance bound properties.

5.3 Simulation Design

We assess the performance of our proposed test procedure by conducting a Monte Carlo experiment. A detailed documentation of the framework of the experiment that follows can be obtained from Patton and Timmermann [2012].

Let the target variable y_t evolve according to an AR(1) process

$$y_t = \mu_y + \phi(y_{t-1} - \mu_y) + \epsilon_t, \quad \varepsilon_t \sim iidN(0, \sigma_\epsilon^2), \quad t = 1, \ldots, 100, \qquad (5.6)$$

with $\phi = 0.5$, $\sigma_y^2 = 0.5$, $\mu_y = 0.75$. The optimal forecast is given by

$$\hat{y}^*_{t|t-h} = E_{t-h}[y_t] = \mu_y + \phi^h(y_{t-h} - \mu_y) \;\; \forall \; h = 1, \ldots, H = 4.$$

The DGP of y_t and its optimal forecast are further contaminated to set the stage for the experiment that follows.

i. Measurement error: owing to the suspicion that the performance of rationality tests that rely on the target variable versus tests that only use forecasts are heavily influenced by measurement errors in the underlying target variable, y_t. The target variable $\tilde{y}_t$ is assumed to be observed with error, ψ_t

$$\tilde{y}_t = y_t + \psi_t, \quad \psi_t \sim iidN(0, \sigma^2_\psi). \tag{5.7}$$

where σ_ψ can take three different values: (i) zero, $\sigma_\psi = 0$; (ii) medium, $\sigma_\psi = \sqrt{0.7}\sigma_y$; and (ii) high, $\sigma_\psi = 2\sqrt{0.7}\sigma_y$.

ii. Suboptimal forecasts: in order to study the power of the optimality tests, it is assumed that the optimal forecast is contaminated by the same level of noise at all horizons:

$$\hat{y}_{t|t-h} = \hat{y}^*_{t|t-h} + \sigma_{\xi,h}\xi_{t,t-h}, \quad \xi_{t,t-h} \sim iidN(0, 1), \tag{5.8}$$

where $\sigma_{\xi,h} = \sqrt{0.7}\sigma_y$ for all h. A simple extension to the case when the standard deviation is increasing in the horizon is also considered. In this case, $\sigma_{\xi,h}$ is assumed to be equal to $\frac{2(h-1)}{7} \times \sqrt{0.7}\sigma_y$ for $h = 1, 2, \ldots, H$.

5.3.1 Theoretical moments

As a basis for checking the performance of our proposed Bayesian inference technique, we compute, from a frequentist perspective, the theoretical moments of the various nonlinear function defining the implications of forecast rationality as discussed in Patton and Timmermann [2012] and check when the implications of forecast rationality are violated.

1. **Monotonicity (Increasing) of Mean Squared Error (MSE)**: Forecast rationality implies that under mean square error, mean squared forecast error should be a weakly increasing function of the forecast horizon. Based on our simulation design,

$$\begin{aligned} E[e^2_{t|t-h}] &= E[(\tilde{y}_t - \hat{y}_{t|t-h})^2] \\ &= \sigma^2_y(1-\phi^{2h}) + \sigma^2_\psi + \sigma^2_{\xi,h}. \end{aligned} \tag{5.9}$$

   Observe that

$$E[e^2_{t|t-h-1}] - E[e^2_{t|t-h}] = \sigma^2_y\phi^{2h}(1-\phi^2) + \sigma^2_{\xi,h+1} - \sigma^2_{\xi,h} \geq 0 \quad \forall h. \tag{5.10}$$

   This shows that MSE is an increasing function of h. Hence, hypothesis tests constructed for checking the monotonicity of the MSE under forecast rationality may not be capable of detecting deviations from optimality.

2. **Monotonicity (Decreasing) of Mean Square Forecast (MSF)**: The internal consistency of a sequence of optimal forecast suggests that the variance of rational forecasts should be a non-increasing function of the forecast horizon.

This implication, under our simulation design, may be checked by evaluating

$$\begin{aligned} E[y^2_{t|t-h}] &= E[(\hat{y}^2_{t|t-h} + \sigma_{\xi,h}\xi_{t,t-h})^2] \\ &= \mu^2_y + \sigma^2_y \phi^{2h} + \sigma^2_{\xi,h}. \end{aligned} \tag{5.11}$$

Here, note that MSF is independent of the magnitude of the measurement error. Two cases can be distinguished here,

Case 1: Constant $\sigma^2_{\xi,h}$: In this case, MSF decreases as h increases. This suggests that the hypothesis tests based on the monotonicity of MSF may be incapable of detecting deviations from optimality.

Case 2: Increasing $\sigma^2_{\xi,h}$: In this case, Figure (5.1) shows an increasing trend in MSF as h increases. This trend contradicts the decreasing trend expected of MSF. Hence, deviations from optimality may be detected with the use of MSF monotonicity test.

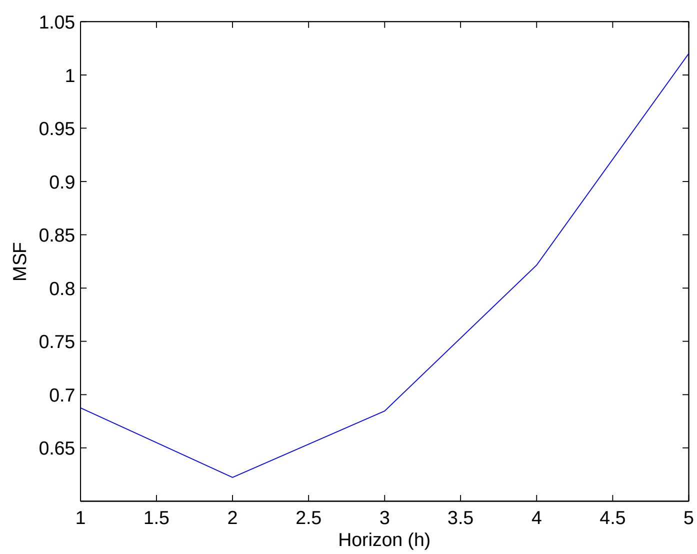


Figure 5.1: Mean squared forecast with increasing noise

3. **Monotonicity of the covariance between forecast and target variable**: Another implication of forecast rationality under mean squared error is that the

covariance between the forecast and its target variable should be a decreasing as the forecast horizon increases. Under of simulation design we have,

$$E[\tilde{y}_t \hat{y}_{t|t-h}] = \mu_y^2 + \phi^{2h}\sigma_y^2. \tag{5.12}$$

The above quantity decreases as h increases. This suggests that the inequality test based on the monotonicity of the covariance between forecast and target variable may be incapable of detecting deviations from optimality.

4. **Covariance bound**: Forecast rationality implies that the variance of forecast revision is bounded above by the covariance between forecast revision and the target variable. This implication will be checked, under our simulation design, by evaluating the following:

$$E[2\tilde{y}_t d_{t|h,h+1} - d^2_{t|h,h+1}] = \phi^{2h}(1-\phi^2)\sigma_y^2 - \sigma^2_{\xi,h} - \sigma^2_{\xi,h+1} \tag{5.13}$$

Case 1 Constant $\sigma_{\xi,h}$: The quantity above is a decreasing function of h and at some point it will become negative. This indicates that the COVBOUND test is capable of detecting deviations from optimality.

Case 2 Increasing $\sigma_{\xi,h}$: Figure (5.2) describe the quantity in Equation (5.13). The quantity is mostly negative which implies that test based on Covariance bound is capable of detecting violation of forecast rationality.

5. **Monotonicity (increasing) of Mean Squared Forecast Revision (MSFR)**: Forecast rationality also implies that under mean square error, mean squared forecast revision should be a non decreasing function of the

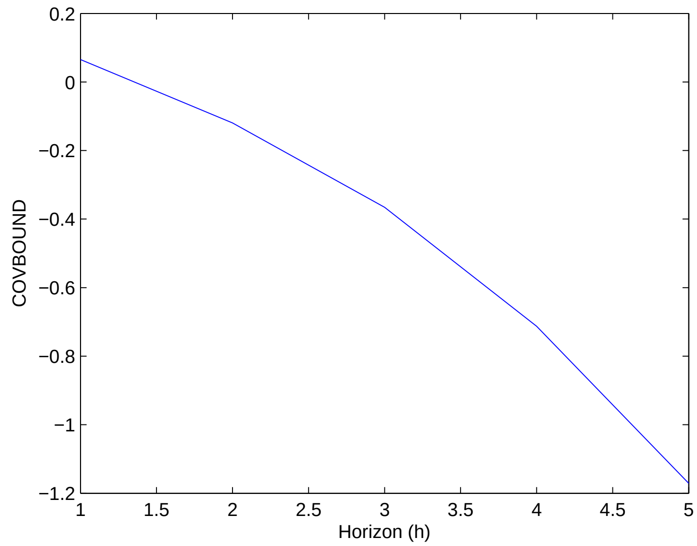


Figure 5.2: Covariance between forecast revision and actual with increasing noise

forecast horizon. This suggests that under our simulation design,

$$\begin{aligned} E[d^2_{t|1,h}] &= E[(\hat{y}_{t|t-1} - \hat{y}_{t|t-h})^2] \\ &= \sigma^2_y(\phi^2 - \phi^{2h}) + \sigma^2_{\xi,1} + \sigma^2_{\xi,h}, \end{aligned} \tag{5.14}$$

MSFR is an increasing function of h i.e. tests of the monotonicity property of MSFR may not be capable of detecting deviations from forecast rationality under our simulation design.

6. **Monotonicity of Covariance between long horizon forecast and target (proxy)**: In the absence of the target variable, forecast rationality suggests that the covariance with the shortest horizon forecast, which is used as a proxy for the unavailable target variable, and longer horizon forecast should be a weakly decreasing function of the forecast horizon h. In our simulation design

$$E[\hat{y}_{t|t-1}\hat{y}_{t|t-h}] = \mu^2_y + \phi^{2h}\sigma^2_y \tag{5.15}$$

similar deductions made in item (3) is arrived at here.

7. **Covariance bound with proxy**: Replacing the target variable with a proxy, shortest horizon forecast, forecast rationality implies that the variance of forecast revision using the proxy is bounded above by the covariance between the modified forecast revision and the proxy. Assessing this implication in our simulation design

$$E[2\hat{y}_{t|t-1}d_{t|h,h+1} - d^2_{t|h,h+1}] = \phi^{2h}(1-\phi^2)\sigma^2_y - \sigma^2_{\xi,h} - \sigma^2_{\xi,h+1} \tag{5.16}$$

gives a similar result as obtained in item (4).

In summary, from the analytical results of the theoretical moments described above, we can deduce that only covariance bound tests are capable of detecting deviations from optimality when the standard deviation, $\sigma_{\xi,h}$, of the noise in the optimal forecast is kept constant over all horizons. On the other hand, the mean squared forecast (MSF) and covariance bound tests should be able to detect deviations from optimality if the standard deviation of the noise in the optimal forecast is allowed to increase with the forecast horizon.

5.3.2 Results from Monte Carlo Simulation

Let the random vector $Y_t = (y_t, y_{t|t-1}, y_{t|t-1}, \ldots, y_{t|t-H})'$ representing the observation at time t follow a multivariate normal distribution i.e.

$$Y_t \sim \mathcal{N}_{H+1}(\mu, \Sigma) \tag{5.17}$$

for $t = 1, 2, \ldots, T$, with

$$\mu = \begin{pmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_H \end{pmatrix}, \Sigma = \begin{pmatrix} \sigma_0^2 & \sigma_{0,1} & \cdots & \sigma_{0,H} \\ \sigma_{1,0} & \sigma_1^2 & \cdots & \sigma_{1,H} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{H,0} & \sigma_{H,1} & \cdots & \sigma_H^2 \end{pmatrix}.$$

In the unconstrained case, we combine a non-informative prior on the parameters

$$\pi(\mu, \Sigma^{-1}) \propto |\Sigma^{-1}|^{-\frac{H+2}{2}}, \tag{5.18}$$

with the likelihood function

$$\mathcal{L}(\mu, \Sigma^{-1}|Y_1', \ldots, Y_T) \propto |\Sigma^{-1}|^{T/2} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}(Y_t - \mu)'\Sigma^{-1}(Y_t - \mu)\right), \tag{5.19}$$

to obtain the joint posterior distribution

$$f_u(\mu, \Sigma^{-1}|Y_1', \ldots, Y_T) \propto |\Sigma^{-1}|^{\frac{T-(H+2)}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}(Y_t - \mu)'\Sigma^{-1}(Y_t - \mu)\right). \tag{5.20}$$

Whereas, in the constrained case we set the prior on the parameters to be the product of the quantity in Equation (5.18) and an indicator function accounting for the inequality. Combining this modified prior with the likelihood function in Equation (5.19) leads to the following joint posterior probability for the constrained model

$$f_c(\mu, \Sigma^{-1}|Y_1', \ldots, Y_T) \propto |\Sigma^{-1}|^{\frac{T-(H+2)}{2}} \exp\left(-\frac{1}{2}\sum_{t=1}^{T}(Y_t - \mu)'\Sigma^{-1}(Y_t - \mu)\right)\mathbb{I}_{\theta\in Q}, \tag{5.21}$$

where $\theta = (\mu, \Sigma)$ and Q represents the region for which the inequality constraints are satisfied. These inequalities are derived from the implications of forecast rationality (See Section (5.3.1) for summary) and are summarized as hypothesis in the following:

1. Monotonicity (Increasing) of Mean Squared Error (MSE).

   Let the forecast error be

$$e_{t|t-h} = \begin{cases} y_t - y_{t|t-1} = (1,-1,0,\ldots,0)Y_t, & \text{for } h=1, \\ y_t - y_{t|t-h} = (1,0,\ldots,0,-1,0,\ldots,0)Y_t, & \text{for } h=2,\ldots,H-1 \\ y_t - y_{t|t-H} = (1,0,\ldots,0,-1)Y_t & \text{for } h=H, \end{cases} \tag{5.22}$$

   and the second non-central moments

$$E[e^2_{t|t-h}] = \begin{cases} (1,-1,0,\ldots,0)\Sigma(1,-1,0,\ldots,0)' + ((1,-1,0,\ldots,0)\mu)^2, & \text{for } h=1, \\ (1,0,\ldots,0,-1,0,\ldots,0)\Sigma(1,0,\ldots,0,-1,0,\ldots,0)' & \\ +((1,0,\ldots,0,-1,0,\ldots,0)\mu)^2, & \text{for } h=2,\ldots,H-1 \\ (1,0,\ldots,0,-1)\Sigma(1,0,\ldots,0,-1)' + ((1,0,\ldots,0,-1)\mu)^2 & \text{for } h=H, \end{cases} \tag{5.23}$$

   then the Null hypothesis of forecast rationality we have

$$H_0 : E[e^2_{t|t-h-1}] - E[e^2_{t|t-h}] \geq 0 \quad \forall \;\; h = 1,2,\ldots,H-1 \tag{5.24}$$

2. Monotonicity (Decreasing) of Mean Square Forecast (MSF).

   Let the forecast be represented by

$$y_{t|t-h} = \begin{cases} (0,1,0,\ldots,0)Y_t, & \text{for } h=1, \\ (0,\ldots,0,1,0,\ldots,0)Y_t, & \text{for } h=2,\ldots,H-1 \\ (0,0,\ldots,0,1)Y_t & \text{for } h=H, \end{cases} \tag{5.25}$$

and the second non-central moments

$$E[y_{t|t-h}^2] = \begin{cases} (0,1,0,\ldots,0)\Sigma(0,1,0,\ldots,0)' + ((0,1,0,\ldots,0)\mu)^2, & \text{for } h=1, \\ (0,\ldots,0,1,0,\ldots,0)\Sigma(0,\ldots,0,1,0,\ldots,0)' & \\ +((0,\ldots,0,1,0,\ldots,0)\mu)^2, & \text{for } h=2,\ldots,H-1 \\ (0,\ldots,0,1)\Sigma(0,\ldots,0,1)' + ((0,\ldots,0,1)\mu)^2 & \text{for } h=H, \end{cases} \tag{5.26}$$

then the Null hypothesis of forecast rationality is

$$H_0 : E[y_{t|t-h-1}^2] - E[y_{t|t-h}^2] \leq 0 \quad \forall \ h = 1,2,\ldots,H-1$$

3. Monotonicity of Covariance Between the Forecast and Target Variable.

$$E[y_t y_{t|t-h}] = \begin{cases} (1,0,\ldots,0)\Sigma(0,1,0,\ldots,0)' + (1,0,\ldots,0)\mu\mu'(0,1,0,\ldots,0)', & \text{for } h=1, \\ (1,0,\ldots,0)\Sigma(0,\ldots,0,1,0,\ldots,0)' & \\ +(1,0,\ldots,0)\mu\mu'(0,\ldots,0,1,0,\ldots,0)', & \text{for } h=2,\ldots,H-1 \\ (1,0,\ldots,0)\Sigma(0,\ldots,0,1)' + (1,0,\ldots,0)\mu\mu'(0,\ldots,0,1)' & \text{for } h=H, \end{cases} \tag{5.27}$$

Null hypothesis:

$$H_0 : E[y_t y_{t|t-h-1}] - E[y_t y_{t|t-h}] \leq 0 \quad \forall \ h = 1,2,\ldots,H-1$$

4. Bounds on Variances of Forecast Revisions.

Let forecast revision be defined by

$$d_{t|h,h+1} = y_{t|t-h} - y_{t|t-h-l} = \begin{cases} (0,1,-1,0,\dots,0)Y_t \ \text{ for } \ h=1 \\ (0,\dots,0,1,-1,0,\dots,0)Y_t \ \text{ for } \ h=2,\dots,H-2 \\ (0,\dots,0,1,-1)Y_t \ \text{ for } \ h=H-1 \end{cases} \tag{5.28}$$

and the second non-central moments

$$E[d^2_{t|h,h+1}] = \begin{cases} (0,1,-1,0,\dots,0)\Sigma(0,1,-1,0,\dots,0)' + ((0,1,-1,0,\dots,0)\mu)^2, \quad \text{for } \ h=1, \\ (0,\dots,0,1,-1,0,\dots,0)\Sigma(0,\dots,0,1,-1,0,\dots,0)' \\ +((0,\dots,0,1,-1,0,\dots,0)\mu)^2, \qquad \text{for } \ h=2,\dots,H-1 \\ (0,\dots,0,1,-1)\Sigma(0,\dots,0,1,-1)' + ((0,\dots,0,1,-1)\mu)^2 \ \text{ for } \ h=H-1, \end{cases} \tag{5.29}$$

then Null hypothesis:

$$H_0 : E[2y_t d_{t|h,h+1} - d^2_{t|h,h+1}] \geq 0 \ \text{ for } \quad h=1,\dots,H-1.$$

5. Mean Squared Forecast Revision (MSFR).

   Null hypothesis:

$$H_0 : E[d^2_{t|1,h+2} - d^2_{t|1,h+1}] \geq 0 \quad \text{for } \ h=1,\dots,H-2$$

6. Monotonicity of Covariance Between the Forecast and Target(Proxy) Variable.

   Null hypothesis:

$$H_0 : E[y_{t|t-1} y_{t|t-h-2}] - E[y_{t|t-1} y_{t|t-h-1}] \leq 0 \quad \text{for } \ h=1,\dots,H-2$$

7. Bounds on Covariances of Forecast Revisions with Proxy:

   Null hypothesis:

$$H_0 : E[2y_{t|t-1}d_{t|h,h+1} - d^2_{t|h,h+1}] \geq 0 \quad \text{for} \quad h = 2, \ldots, H-1$$

These hypothesis will be evaluated according to our proposed Bayesian inference procedure. Samples of (μ, Σ) are generated from the unconstrained posterior distribution, Equation (5.20), via Gibbs sampling scheme. We draw samples of μ from its full conditional unconstrained posterior distribution given Σ and observation $Y_1, \ldots, Y_T$ i.e.

$$f(\mu|Y_1', \ldots, Y_T', \Sigma) \propto |\Sigma_\mu|^{-1} \exp\left(-\frac{1}{2}(\mu - \tau_\mu)'\Sigma_\mu^{-1}(\mu - \tau_\mu)\right) \tag{5.30}$$

where

$$\Sigma_\mu = \frac{\Sigma}{T}, \quad \text{and} \quad \tau_\mu = \Sigma_\mu \Sigma^{-1} \sum_{t=1}^{T} Y_t = T^{-1} \sum_{t=1}^{T} Y_t.$$

Σ on the other hand, will be generated from its full conditional unconstrained posterior distribution given μ and observation $Y_1, \ldots, Y_T$

$$f(\Sigma^{-1}|Y_1, \ldots, Y_T, \mu) = Wishart(\nu_\Sigma, R_\Sigma), \tag{5.31}$$

with

$$\nu_\Sigma = T, \quad \text{and} \quad R_\Sigma = \left(\sum_{t=1}^{T}(Y_t - \mu)(Y_t - \mu)'\right)^{-1}.$$

Haven generated samples of (μ, Σ) via the above Gibbs sampling technique, we evaluate the posterior odds ratio in favour of the constrained model using Equation (5.5). The results presented in Table (5.1) and (5.2) below are based on experiment conducted on 1000 simulated datasets each consisting of 100 observation. The largest horizon, H, in this experiment is 4. The Bayesian exercise is conducted using 10000 samples from the posterior distribution. To reduce serial correlation in our draws, we consider every 10th draw of the Gibbs iteration in our Bayesian exercise.

Table 5.1: Constraints Probability $\hat{p}_{1|2}$ in the presence of measurement error

Meas. error std.:	Zero $\sigma_\psi = 0$	Med $\sigma_\psi = \sqrt{0.7}\sigma_y$	High $\sigma_\psi = 2\sqrt{0.7}\sigma_y$
Test			
Increasing MSE	0.0394	0.0320	0.0236
	(0.0228)	(0.0213)	(0.0184)
Decreasing COV	0.0493	0.0447	0.0366
	(0.0229)	(0.0230)	(0.0220)
COV BOUND	0.0340	0.0301	0.0241
	(0.0246)	(0.0232)	(0.0203)
Decreasing MSF	0.0442	0.0442	0.0442
	(0.0254)	(0.0255)	(0.0255)
Increasing MSFR	0.0596	0.0595	0.0595
	(0.0187)	(0.0187)	(0.0187)
Decreasing COV,	0.0568	0.0568	0.0568
with proxy	(0.0211)	(0.0211)	(0.0211)
COV BOUND,	0.0420	0.0420	0.0420
with proxy	(0.0250)	(0.0250)	(0.0250)

NOTES: This table presents the outcome of 1,000 Monte Carlo simulations of the constraint probability of various forecast optimality tests. Data are generated by a first-order autoregressive process with parameters $\phi = 0.5$, $\sigma_y^2 = 0.5$, $\mu_y = 0.75$. We consider three levels of error in the measured value of the target variable (high, medium, and zero). Optimal forecasts are generated under the assumption that this process (and its parameter values) are known to forecasters. The simulations assume a sample of 100 observations and tests labeled with proxy refer to cases where the one-period forecast is used in place of the predicted variable. The standard deviation of constraint probabilities for each the 1000 Monte Carlo experiment is given in the parenthesis.

The econometric model in our Monte Carlo experiment is a particular case of linear regression models subject to inequality constraints. Within this framework, we defer from most literature on inferential exercise on inequality constraints by allowing the inequality constraints depend on both the scale matrix and the coefficients of the regression model in, possibly, a nonlinear manner. Table 5.1 and 5.2 respectively report the constraint probability, $\hat{p}_{1|2}$, in the presence of measurement error and suboptimality in the forecast. As noted in Section 2, a substantive result of these probabilities can be obtained by evaluating the limiting ratio of prior odds ratio of the each constraint i.e. the probabilities in Table (5.1) and (5.2) need to be scaled appropriately in order to obtain substantive results showing whether the constrained

Table 5.2: Constraints Probability $\hat{p}_{1|2}$ in the presence of suboptimal forecast.

	Equal noise across all forecast horizon			Increasing noise across forecast horizon		
	Zero	Med	High	Zero	Med	High
Meas. error std.:	$\sigma_\psi = 0$	$\sigma_\psi = \sqrt{0.7}\sigma_y$	$\sigma_\psi = 2\sqrt{0.7}\sigma_y$	$\sigma_\psi = 0$	$\sigma_\psi = \sqrt{0.7}\sigma_y$	$\sigma_\psi = 2\sqrt{0.7}\sigma_y$
Test						
Increasing MSE	0.0079(0.0085)	0.0068(0.0077)	0.0058(0.0071)	0.0607(0.0150)	0.0481(0.0175)	0.0314(0.0179)
Decreasing COV	0.0109(.01000)	0.0098(0.0097)	0.0082(0.0089)	0.0216(0.0161)	0.0198(0.0156)	0.0167(0.0145)
COV BOUND	0.0000(0.0000)	0.0000(0.0000)	0.0000(0.0000)	0.0000(0.0003)	0.0001(0.0008)	0.0004(0.0016)
Decreasing MSF	0.0080(0.0089)	0.0080(0.0089)	0.0080(0.0089)	0.0006(0.0019)	0.0006(0.0020)	0.0006(0.0020)
Increasing MSFR	0.0156(0.0147)	0.0156(0.0147)	0.0156(0.0147)	0.0771(0.0057)	0.0771(0.0057)	0.0771(0.0057)
Decreasing COV,with proxy	0.0175(0.0152)	0.0175(0.0152)	0.0175(0.0152)	0.0245(0.0175)	0.0244(0.0175)	0.0244(0.0175)
COV BOUND,with proxy	0.0000(0.0000)	0.0000(0.0000)	0.0000(0.0000)	0.0000(0.0000)	0.0000(0.0000)	0.0000(0.0000)

NOTES: This table presents the outcome of 1,000 Monte Carlo simulations of the constraint probability of various forecast optimality tests. Data are generated by a first-order autoregressive process with parameters $\phi = 0.5$, $\sigma_y^2 = 0.5$, $\mu_y = 0.75$. We consider three levels of error in the measured value of the target variable (high, medium, and zero). Optimal forecasts are generated under the assumption that this process (and its parameter values) are known to forecasters. Constraint probability is studied against suboptimal forecasts obtained when forecasts are contaminated by the same level of noise across all horizons and when forecasts are contaminated by noise that increases in the horizon. The simulations assume a sample of 100 observations and tests labeled with proxy refer to cases where the one-period forecast is used in place of the predicted variable. The standard deviation of constraint probabilities for each the 1000 Monte Carlo experiment is given in the parenthesis.

model is competitive with the unconstrained model or not. Evidently, irrespective of the scaling factor used, in Table (5.2) covariance bounds tests confirms our deductions from the theoretical moments i.e. the forecast are not sequentially rational since they violates the covariance bound implication of forecast rationality.

5.4 Empirical Application

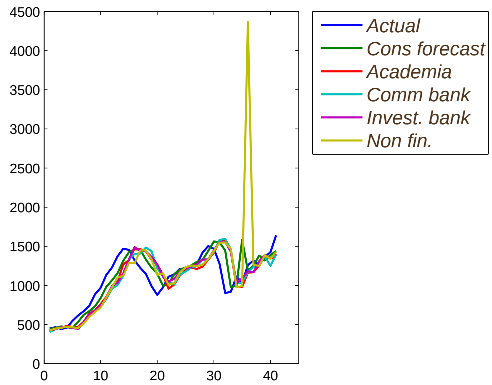
Our database consists of S&P 500 price forecasts obtained from the Livingston Surveys of Professional Forecasters. The Livingston Survey is maintained by the Federal Reserve Bank of Philadelphia and is available from June 1957 to June 2013. The survey data contain forecasts of S&P 500 stock market index at four horizons: 6 months, 12 months, 18 months and 24 months. The survey also classifies forecasters into groups: Academic institutions, Commercial banks, Investment banks, Non-Financial business among others. We limit attention to the sample period June 1993 to June 2013 and two forecast horizons 6 and 12 months for two main reasons. Firstly, in 1992, the Federal Reserve Bank of Philadelphia introduced important changes to the survey design and the handling of the data. Many of these changes were made to overcome well-known obstacles to the academic analysis of the survey responses. One important methodological flaw addressed is survey is the absence of a consistent basic index for computing expected market returns. Prior to this period, researchers make ad hoc assumptions about the dates on which the survey participants filled in the questionnaire in order to compute the expected returns on the stock market index (see Lakonishok [1980] and Pearce [1984]). This approach, as pointed out by Dokko and Edelstein [1989], leads to diverging conclusion made by researchers. The correction made is the questionnaire was to ask forecasters to forecast the current December value for the index (as well as June and the following December), so that there are definite six- and 12-month forecasts. Using the new forecasts avoids this complication stated earlier, as we can now accurately calculate the expected returns semi-annualized expected market returns through:

$$y_{t|t-h} = \left(\left(\frac{SPI_{t|t-h}}{SPI_t} \right)^{(12/h)} - 1 \right) \times 100 \qquad \text{for} \quad h = 6, 12,$$

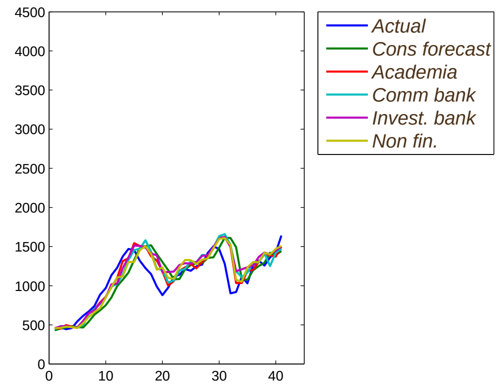
where $y_{t|t-h}$, is the annualized S&P 500 return forecast of at time t for the period starting from the survey publication date $(t-h)$ to h months in the future. Similarly, $SPI_{t|t-h}$ is the S&P 500 stock price forecast (averaged across the forecasters) for h months in the future and SPI_t is the stock price forecast (averaged across the all forecasters or forecasters in a sector) for the surveys publication date. For the same period, S&P 500 price data were obtained from yahoo finance (http://www.finance.yahoo.com.) in order to calculate the realized annualized rate of return.

The second issue relates to the number of available multi-horizon forecasts. An important requirement in our proposed rationality test relies on the availability of multi-horizon forecasts. This requirement provides a means for studying the internal consistency in term structure of the multi-horizon forecast. Unfortunately, we observed that majority of participants in the survey only report forecast for two horizons i.e. 6 and 12-months horizons. Also, we notice that the number respondents to the survey change over time. These empirical observations provide basis for the choice of our sample period and horizon.

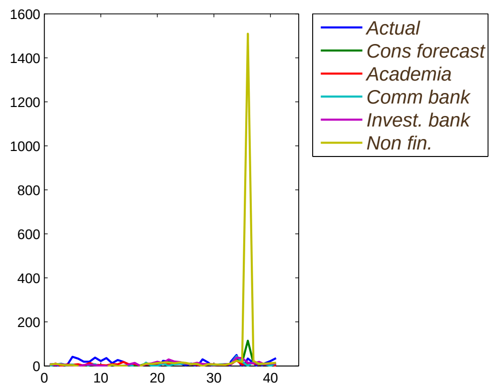
In our empirical exercise, we consider the average responses from the survey and in order to have better understanding of our results we study individual survey responses based on group affiliation. Figure ?? displays the actual, the consensus forecast over all forecasters and the consensus forecast for each forecaster category for both the stock market index and their corresponding returns. Except from the large deviation from the target series produced by the non financial business at the later part of the series, all forecast produced by the forecasters are observed to be close to the actual values. Table 5.3 reports the posterior odds probabilities for comparing the constrained model against the the unconstrained model. From the



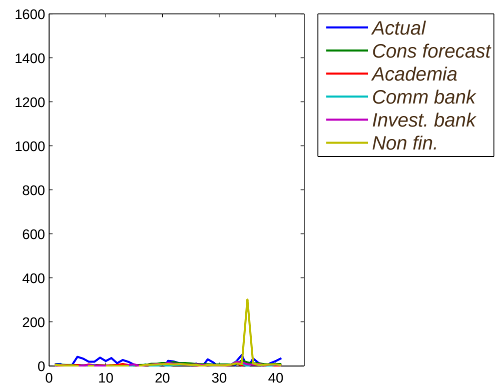
(a) 6 months forecast of S&P 500 index



(b) 12 months forecast of S&P 500 index



(c) 6 months forecast returns of S&P 500



(d) 12 months forecast returns of S&P 500

Figure 5.3: Forecast of the S&P 500 index and returns from June 1993 to June 2013.

Table 5.3: Constraints Probability $\hat{p}_{1|2}$ in the presence of measurement error

	Full Sample	Academic Institutions	Commercial Banks	Investment Banks	Non-Financial Business
Test					
Increasing MSE	0.0092	0.0895	0.0507	0.0084	0
Decreasing COV	0.0990	0.0990	0.0990	0.0990	0.0990
COV BOUND	0.0906	0.0980	0.0880	0.0724	0.0922
Decreasing MSF	0.0192	0.0924	0.0494	0.0273	0

Bibliography

Kevin Aretz, Söhnke M Bartram, and Peter F Pope. Asymmetric loss functions and the rationality of expected stock returns. *International Journal of Forecasting*, 27(2):413–437, 2011. URL http://www.sciencedirect.com/science/article/pii/S0169207009001769.

C. Capistran. Optimality Tests for Multi-Horizon Forecast. Working Paper 2007-14, Banco de Mexico, 2007.

G. Chamberlain and E. E. Leamer. Matrix weighted averages and posterior bounds. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38(1):73–84, 1976.

W. W. Davis. Bayesian analysis of the linear model subject to linear inequality constraints. *Journal of the American Statistical Association*, 73(363):573–579, 1978. URL http://amstat.tandfonline.com/doi/full/10.1080/01621459.1978.10480057.

Werner FM De Bondt. What do economists know about the stock market. *Journal of Portfolio Management*, 17(2):84–91, 1991.

F. X. Diebold and J. Lopez. Forecast Evaluation and Combination. In Elsevier, editor, *Hanbook of Stattistics*, pages 241 – 268. Amsterdam North, 1996.

Yoon Dokko and Robert H Edelstein. How well do economists forecast stock market prices? a study of the livingston surveys. *The American Economic Review*, 79(4): 865–871, 1989. URL http://www.jstor.org/stable/10.2307/1827938.

G. Elliot and A. Timmermann. Optimal Forecast Combination Under General Loss Function and Forecast Error Distributions. *Journal of Econometric*, 122:47 – 79, 2004.

G. Elliot, I. Komunjer, and A. Timmermann. Estimation and Testing of Forecast Rationality under Flexible Loss. *Review of Economic Studies*, 72:1107 – 1125, 2005.

J. Geweke. Exact inference in the inequality constrained normal linear regression model. *Journal of Applied Econometrics*, 1(2):127–141, 1986. URL http://onlinelibrary.wiley.com/doi/10.1002/jae.3950010203/abstract.

J. Geweke and C. Whiteman. Bayesian Forecasting. In G. Elliott, C.W.J. Granger, and A. Timmermann, editors, *Handbook of Economic Forecasting*, pages 3 – 80. Elsevier, 2006.

J. F Geweke. Bayesian Inference for Linear Models Subject to Linear inequality Constraints. Federal Reserve Bank of Mineapolis Research Department Working paper, 1995.

R. Giacomini and B. Rossi. Forecast Comparisons in Unstable Environments. *Journal of Applied Econometrics*, 25(4):595–620, 2010. URL http://onlinelibrary.wiley.com/doi/10.1002/jae.1177/full.

C. Gourieroux, A. Holly, and A. Monfort. Likelihood ratio test, Wald test, and Kuhn-Tucker test in linear models with inequality constraints on the regression parameters. *Econometrica: Journal of the Econometric Society*, 50:63–80, 1982. URL http://www.jstor.org/stable/10.2307/1912529.

C.W.J. Granger and P.A. Newbold. *Forecasting Economic Time Series*. Economic Theory, Econometrics and Mathematical Economics Series. Academic Press, 1986. ISBN 9780122951848. URL http://books.google.it/books?id=xGq9QgAACAAJ.

P. R. Hansen. A test for superior predictive ability. *Journal of Business & Economic Statistics*, 23(4):365–380, 2005. URL http://amstat.tandfonline.com/doi/abs/10.1198/073500105000000063.

M Hashem Pesaran, Andreas Pick, and Mikhail Pranovich. Optimal forecasts in the presence of structural breaks. *Journal of Econometrics*, 2013. URL http://www.sciencedirect.com/science/article/pii/S0304407613000687.

L. Hoogerheide, F. Ravazzolo, and H. K. Van Dijk. Basktesting Value-at-Risk Using Forecasts for Multiple Horizons, a Comment on the Forecast Rationality Tests Based on Multi-Horizon Bounds of A.J. Patton and A. Timmermann. *Journal of Business & Economic Statistics*, 30:32 – 35, 2012.

Ivana Komunjer and Michael T Owyang. Multivariate forecast evaluation and rationality testing. *Review of Economics and Statistics*, 94(4):1066–1080, 2012.

Josef Lakonishok. Stock market return expectations: Some general properties. *The Journal of Finance*, 35(4):921–931, 1980. URL http://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1980.tb03510.x/abstract.

J. Mincer and V. Zarnowitz. The Evaluation of Economic forecasts. In *Economic Forecasts and Expectations: Analysis of Forecasting Behavior and Performance*, 1969.

Franco Modigliani and Robert J Shiller. Inflation, rational expectations and the term structure of interest rates. *Economica*, 40(157):12–43, 1973. URL http://www.jstor.org/stable/10.2307/2552679.

John F Muth. Rational expectations and the theory of price movements. *Econometrica: Journal of the Econometric Society*, pages 315–335, 1961. URL http://www.jstor.org/stable/10.2307/1909635.

Charles R Nelson. Rational expectations and the estimation of econometric models. *International Economic Review*, 16(3):555–561, 1975. URL http://www.jstor.org/stable/10.2307/2525996.

A. Patton and A. Timmermann. Testing Forecast Optimality Under Unknown Loss. *Journal of the American Statistical Association*, 102:1172 – 1184, 2007a.

A. J. Patton and A. Timmermann. Properties of Optimal Forecasts under Asymmetric Loss and Nonlinearity. *Journal of Econometric*, 140:884 – 918, 2007b.

A.J. Patton and A. Timmermann. Rejoinder. *Journal of Business & Economic Statistics*, 30(1):1–17, 2012.

Douglas K Pearce. An empirical analysis of expected stock price movements. *Journal of Money, Credit and Banking*, 16(3):317–327, 1984. URL http://www.jstor.org/stable/10.2307/1992219.

Christian Pierdzioch and Jan-Christoph Rülke. Forecasting stock prices: Do forecasters herd? *Economics Letters*, 116(3):326–329, 2012. URL http://www.sciencedirect.com/science/article/pii/S0165176512001152.

F. Ross. Survey of Economic Forecasting Techniques: A Survey Article. *Econometrica*, 23:363 – 395, 1955.

B. Rossi. Comment to: Forecast Rationality Tests Based on Multi-Horizon Bounds, by Patton, A. and A. Timmermann. *Journal of Business & Economic Statistics*, 30(1):25 –, 2012.

B. Rossi and T. Sekhposyan. Forecast Optimality Tests in the Presence of Instabilities. *Economic Research Initiatives at Duke (ERID) Working Paper*, (109), 2011.

Thomas J Sargent. Rational expectations and the term structure of interest rates. *Journal of Money, Credit and Banking*, 4(1):74–97, 1972. URL http://www.jstor.org/stable/10.2307/1991403.

P. Söderlind. Predicting stock price movements: regressions versus economists. *Applied Economics Letters*, 17(9):869–874, 2010. doi: 10.1080/17446540802584871. URL http://www.tandfonline.com/doi/abs/10.1080/17446540802584871.

H. White. A reality check for data snooping. *Econometrica*, 68(5):1097–1126, 2003. URL http://onlinelibrary.wiley.com/doi/10.1111/1468-0262.00152/abstract.

F. A. Wolak. An exact test for multiple inequality and equality constraints in the linear regression model. *Journal of the American Statistical Association*, 82(399):782–793, 1987. URL http://amstat.tandfonline.com/doi/abs/10.1080/01621459.1987.10478499.

Frank A Wolak. Testing inequality constraints in linear econometric models. *Journal of econometrics*, 41(2):205–235, 1989. URL http://www.sciencedirect.com/science/article/pii/0304407689900948.

A. Zellner. Bayesian Estimation and Prediction Using Asymmetric Loss Functions. *Journal of the American Statistical Association*, 81:446 – 451, 1986.

Arnold Zellner. *An introduction to Bayesian inference in econometrics*. New York : John Wiley & Sons, 1996.

Estratto per riassunto della tesi di dottorato

Studente: OSUNTUYI AYOKUNLE ANTHONY
Matricola: 955710
Dottorato: Economia
Ciclo: 25

Titlo della tesi: Essays on Bayesian inference with financial applications

Abstract: This thesis is composed of two main research lines. The first line, developed in chapters 2 to 4, deals with frequentist and Bayesian estimation of regime-switching GARCH models and its application to risk management on energy markets, while the second part, which corresponds to chapter 5, focuses on forecast rationality testing within a Bayesian framework.

Chapter 2 presents a unified mathematical framework for characterizing the class of MSGARCH models based on collapsing the regimes in order to eliminate the usual path dependence problem. Within this framework, two new models (identified as Basic model and Simplified Klaassen model) are proposed as alternative specifications of the MS-GARCH model. Using Maximum Likelihood Estimation, we estimate the parameters of the different models within this family and compare their performance on both simulation and empirical exercises. Chapter 3 proposes new efficient Monte Carlo simulation techniques based on multiple proposal Metropolis. The application to approximated inference for regime-switching GARCH models is there discussed. In Chapter 4, we provide an extension of our efficient Monte Carlo simulation algorithm to a multi-chain Markov switching multivariate GARCH model and apply it to risk management in commodity market. More specifically we focus on futures commodity market and suggest a dynamic and robust minimum variance hedging strategy which accounts for model parameter uncertainty. In chapter 5, we propose a new Bayesian inference procedure for testing the monotonicity properties of second moment bounds across several horizons presented in Patton and Timmermann [2012].

Estratto: Questa tesi sviluppa due principali linee di ricerca. La prima, presentata nei capitoli 2,3 e 4, tratta l'inferenza frequentista e classica per i modelli GARCH a cambiamento di regime (MSGARCH) e la loro applicazione alla gestione del rischio nel mercato energetico, mentre la seconda linea, presentata nel capitolo 5, si concentra sul verifica statistica della razionalità previsiva proponendo un approccio bayesiano.

Il capitolo 2 presenta un approccio matematico all'approssimazione della dinamica MSGARCH che si fonda sul collasso dei regimi e che consente di eliminare il problema della dipendenza del modello MSGARCH dalla traiettoria passata degli stati. Applicando questo metodo di approssimazione sono stati identificati due nuovi modelli approssimati, chiamati modello di base e modello di Klaassen semplificato, che sono stati poi utilizzati nelle procedure di inferenza per modelli MSGARCH in alternativa alle approssimazioni già esistenti in letteratura. Utilizzando il metodo di stima di massima verosimiglianza, si stimano i parametri dei diversi modelli di questa classe e si confronta l'efficienza della stima sia attraverso esempi in simulazioni che

con applicazioni a dati reali. Il capitolo 3 propone una nuova ed efficiente tecnica di simulazione Monte Carlo basata su algoritmi Metropolis con distribuzioni strumentali multiple. Nello stesso capitolo si discute anche dell'applicazione dei metodi di simulazione all'inferenza approssimata per modelli MSGARCH. Il capitolo 4 presenta una applicazione dei nuovi metodi di simulazione Monte Carlo per modelli MSGCARCH bayesiani, sviluppati nei precedenti capitoli della tesi, alla gestione del rischio nei mercati delle commodity. In modo più specifico, il capitolo si concentra su sui contratti a termine nel mercato delle commodity e suggerisce una strategia di copertura dinamica, robusta e a varianza minima, che include nella decisione ottima l'elemento di incertezza sui parametri del modello. Nel capitolo 5 si propone una nuova procedura di inferenza bayesiana per verificare statisticamente le proprietà di monotonia del limite sui momenti secondi su più orizzonti temporali che sono state discusse in Patton and Timmermann [2012].