SYRTO

Systemic Risk Tomography
Signals, Measurements, Transmission channels and Policy Interventions

Markov Switching GARCH models for Bayesian Hedging on Energy Futures Markets

Monica Billio, Roberto Casarin, Anthony Osuntuyi

SYRTO WORKING PAPER SERIES
Working paper n. 12 | 2013

This project is funded by the European Union under the 7th Framework Programme (FP7-SSH/2007-2013) Grant Agreement n° 320270

Markov Switching GARCH models for Bayesian Hedging on Energy Futures Markets

Monica Billio[a,*], Roberto Casarin[a], Anthony Osuntuyi[a,b]

[a]*Department of Economics, University Ca' Foscari of Venice, Fondamenta San Giobbe 873, 30121, Venice, Italy*
[b]*Department of Mathematics, Obafemi Awolowo University, 220005, Ile-Ife, Nigeria*

Abstract

A new Bayesian multi-chain Markov Switching GARCH model for dynamic hedging in energy futures markets is developed: a system of simultaneous equations for return dynamics on the hedged portfolio and futures is introduced. More specifically, both the mean and variance of the hedged portfolio are assumed to be 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 a MS-GARCH model. This formulation has two main practical and conceptual advantages. First, the different states of the discrete processes can be identified as different volatility regimes. Secondly, the parameters can be easily interpreted as different hedging components. Our formulation also provides an avenue to analyze the contribution of the volatility dynamics and state probabilities to the optimal hedge ratio at each point in time. The application of expected utility framework combined with regime-switching models to define a robust minimum variance hedging strategy which accounts for parameter uncertainty is also presented. Evidence of changes in the optimal hedging strategies before and after the financial crisis is found when the proposed robust hedging strategy is applied to crude oil spot and futures markets.

Keywords: Markov-switching, Hedge ratio, Energy futures, GARCH

1. Introduction

Hedging is an investment position taken to mitigate the adverse effect arising from changes in the price of a companion investment. A crucial issue, 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). This paper aims to contribute to this literature in several ways.

First, we propose a new hedging model within the Minimum Variance (MV) approach to hedging. The MV hedging is based on minimizing the risk of a hedged portfolio. This exercise gives the MV hedge ratio defined (see Johnson (1960)) 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 using the estimate of the slope as an 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 the hedging performance, time-varying hedge ratios have been proposed in the literature. Currently in the literature, the estimation of time-varying MV hedge ratios have been developed along two major lines. One approach involves the estimation of the conditional second order moment of the underlying spot and futures returns. The Generalized Autoregressive

*Corresponding author

Email addresses: `billio@unive.it` (Monica Billio), `r.casarin@unive.it` (Roberto Casarin), `ayokunlet@yahoo.com` (Anthony Osuntuyi)

Conditional Heteroscedasticity (GARCH) models have been proposed for this in Haigh and Holt (2002) and Chang et al. (2010), among others. The later approach treats the hedge ratio as a time-varying regression coefficient, and focuses on the estimation of such a parameter (e.g. see Alizadeh and Nomikos (2004), Lee et al. (2006), and Chang et al. (2010)). Note that this hedging strategy works by re-balancing the hedged portfolio on a period-by-period basis. As this may involve huge transaction costs, it may not be worthwhile to use this particular instrument for hedging. Also, it has been well documented in the empirical literature that the class of GARCH models exhibits high persistence of the conditional variance, i.e. the process is close to being nearly integrated. In view of this, some authors allow the optimal hedge ratio to be state-dependent. Alizadeh et al. (2008), Lee and Yoder (2007a), Lee and Yoder (2007b), among others, propose various multivariate Markov-switching (MS) GARCH (MS-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 different characterizations of the time-varying covariance matrix. While Gray's model is attractive, its analytical intractability is a drawback since it cannot be derived using any standard analytical approximation technique.

Our hedging model considers the Billio et al. (2014) approach to MS-GARCH modeling and inference and extends it to bivariate GARCH with multiple and possibly dependent MS processes (multichain MS). More specifically, we assume a system of simultaneous equations modeling both return dynamics on the hedged portfolio and futures. Each component of this system is characterized by a path dependent MS-GARCH process. Our modeling framework is close to the work by Alizadeh and Nomikos (2004), but we differ from Alizadeh and Nomikos (2004) in two ways. The first difference lies in the characterization of the time-varying variance process. While Alizadeh and Nomikos (2004) consider 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 assume that the conditional variances of futures returns is regime independent or that the hidden process characterizing the dynamics of the hedged portfolio is independent of the one influencing the futures returns process. We account for these limitations in our econometric framework. Still regarding the MS-GARCH framework, Sheu and Lee (2012) argue 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 a multichain Markov regime switching GARCH (MCSG) model. In this paper, we also extend the work of Sheu and Lee (2012) by allowing for simultaneous dependence between the Markov chains of the MSCG model.

Another aim of the paper is to develop a robust hedging approach within the MS-GARCH framework. In practice, the parameters in the optimal hedge ratio are unknown, thus optimal hedge ratios are estimated by replacing the unknown parameters by their corresponding estimates. This approach is referred to in the literature as the "plug-in" or Parameter Certainty Equivalent (PCE) principle. Generally speaking, decision makers are left to provide, using an estimation technique of their choice, estimates of the model parameters, and to substitute them directly in the theoretical model. One of the problems with this approach is that it completely ignores estimation risk. Depending on the econometric specification considered for estimating the optimal hedge ratio, large differences are observed in the estimated MV hedge ratios on the same commodity. This observation further suggests that it may be very costly to ignore estimation risk. Another problem relates to the fact that relevant non sample information (such as insider information or subjective prior) available to the hedger are discarded in the decision making process.

We recast the MV hedging model as an expected utility model and deal with the estimation risk problem within this framework. 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 have been pursued by Raiffa and Schlaifer (1961) and DeGroot (2005), among others. A review of the application of this theory to portfolio choice, prior to 1978, is provided in Bawa et al. (1979). A more recent applications 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 analysis, analytical solutions to either the optimization exercise and/or the integration problem are often not achievable. Accordingly, alternative solutions, such as approximation or simulations, are called for. Müller et al. (2004), Müller et al. (2004), among others, proposes simulation-based approaches to the expected utility optimization problem. In this paper, we propose a robust hedging ratio that accounts not only for parameter uncertainty, but also for different states of the market. We follow a Bayesian decision rule (See, for example, Lence and Hayes (1994a) and Lence and Hayes (1994b)) to account for parameter uncertainty in the definition of optimal hedging strategies.

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 present an empirical application of our proposed model to West Texas Intermediate (WTI) crude oil spot and futures prices and compare the result to the conventional OLS method proposed by Ederington (1979). Section 4 concludes the paper and provides suggestions for possible extensions.

2. Bayesian optimal hedging

Let $(Y, \mathcal{Y}, P_\theta)$ be a probability observation space, with $\{P_\theta\}_{\theta\in\Theta}$ 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, correspond to returns on the underlying and returns on the derivative (e.g., option, futures) at time t. Let us define the information set available at time t, as the σ-algebra $\mathcal{F}_t = \sigma(\{\mathbf{y}_s\}_{s\leq t})$ 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.

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

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

where, $E(\cdot|\mathcal{F})$ is the conditional expectation operator, conditioning on a σ-algebra $\mathcal{F}$, $\mathcal{F}_t^\Theta = \sigma(\mathcal{F}_t \vee \mathcal{F}^\Theta)$ the information set generated by the collection of past values of observable process and parameter prior information, $U(\cdot)$ 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 the past values $\mathbf{y}_{1:t-1}$ and the parameter θ. 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 $(RS_t - hRF_t)$. Under these assumptions, the solution of the problem 1 is

$$h_t = \frac{Cov(RS_t, RF_t|\mathcal{F}_{t-1}^\Theta)}{Var(RF_t|\mathcal{F}_{t-1}^\Theta)}. \tag{2}$$

An implicit, and important assumption in (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 the parameter value, θ (parameter uncertainty).

If the hedger does not know the values of θ, the optimal hedge ratio cannot be evaluated since it is a function of θ. The classic 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, (1) becomes

$$\arg\max_{h\in H} E(U|\mathcal{F}_{t-1}) = \arg\max_{h\in H} \int_Y U(r(h, \mathbf{y}_t))p(\mathbf{y}_t|\mathbf{y}_{1:t-1}, \hat{\theta})d\mathbf{y}_t. \tag{3}$$

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 (Lence and Hayes (1994a)). The uncertainty about the parameters in (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.

$$\begin{aligned}
&\arg\max_{h\in H} E(E(U|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1}) = \\
&= \arg\max_{h\in H} \int_{\Theta} \left(\int_{Y} U(r(h,\mathbf{y}_t))p(\mathbf{y}_t|\mathbf{y}_{1:t-1},\theta)d\mathbf{y}_t \right) p(\theta|\mathbf{y}_{1:t-1})d\theta \\
&= \arg\max_{h\in H} \int_{Y} U(r(h,\mathbf{y}))p(\mathbf{y}_t|\mathbf{y}_{1:t-1})d\mathbf{y}_t,
\end{aligned} \tag{4}$$

where $p(\mathbf{y}_t|\mathbf{y}_{1:t-1})$ is the marginal posterior predictive distribution. Unlike (1), (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

$$h_t^{BAY} = \frac{E(Cov(RS_t, RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}{E(Var(RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}. \tag{5}$$

As highlighted in Bawa et al. (1979), applying Bayes' criterion (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 taken into consideration through the posterior distribution in Bayes' method. In contrast, sample information 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. This framework can be further enriched by accounting for ambiguity. See Guidolin and Rinaldi (2013) for a review.

In many situations, obtaining an analytical solution to the Bayesian optimal hedge ratio problem in (4) can be a daunting task. This is because, in some cases, neither the maximum nor the integrals in (4) can be computed analytically, thus demanding alternative approaches such as 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) 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)). It is worth noting that our approach is general and can be applied to several alternative specifications of the utility function (see, for example, Lence and Hayes (1994b) and Haigh and Holt (2002)), other than the quadratic function, exist 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.

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

$$RS_t = \mu + \nu RF_t + \epsilon_t, \qquad \epsilon_t \overset{iid}{\sim} (0, \sigma^2), \tag{6}$$

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 (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 (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 spaces $(X, \mathcal{X})$ and $(Z, \mathcal{Z})$ and unobserved processes, $s_t \in (X, \mathcal{X})$, and $z_t \in (Z, \mathcal{Z})$, $t = 1, \ldots, T$, which represent, respectively, the hedging regime of the portfolio and the volatility state of the 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, $u \leq t$. The following model defines the relationship between the hedged portfolio and the futures market volatility:

$$
\begin{aligned}
RS_t &= \mu(s_t) + \nu(s_t, z_t) RF_t + \sigma_t \eta_t, && \eta_t \overset{iid}{\sim} \mathcal{N}(0,1),\\
\sigma_t^2 &= \gamma(s_t) + \alpha(s_t)\epsilon_{t-1}^2 + \beta(s_t)\sigma_{t-1}^2, &&\\
RF_t &= a(z_t) + \tau_t \zeta_t, && \zeta_t \overset{iid}{\sim} \mathcal{N}(0,1),\\
\tau_t^2 &= \kappa(z_t) + \omega(z_t)\xi_{t-1}^2 + \psi(z_t)\tau_{t-1}^2, &&
\end{aligned}
\tag{7}
$$

where, $\epsilon_t = \sigma_t \eta_t$, $\xi_t = \tau_t \zeta_t$, $\mu(s_t, z_t)$, $\nu(s_t)$, $\gamma(s_t) > 0$, $\alpha(s_t) \geq 0$, $\beta(s_t) \geq 0$, $a(z_t)$, $\kappa(z_t) > 0$, $\omega(z_t) \geq 0$, $\psi(z_t) \geq 0$. As regards the hedging and volatility states, (s_t, z_t), we asume that they take values in the set $\{1, \ldots, M\}^2$, $t = 1, \ldots, T$ and follow a first order Markov chain with transition probabilities

$$
\pi_{ij,kl} = p(s_t = i, z_t = j | s_{t-1} = k, z_{t-1} = l), \quad \sum_{i=1}^{M}\sum_{j=1}^{M} \pi_{ij,kl} = 1 \quad \forall\, k, l = 1, 2, \ldots, M. \tag{8}
$$

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

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

The parameter shift function $\nu(s_t, z_t)$ plays a crucial role in our model, since it allow to separate the contribution of the spot and futures market volatilities to the hedging strategy. We assume:

$$
\nu(s_t, z_t) = \sum_{i,j=1}^{M} \nu_{ij} \mathbb{I}_{s_t=i} \mathbb{I}_{z_t=j}.
$$

In order to simplify the exposition, we define $s_{u:t} = (s_u, \ldots, s_t)$, $z_{u:t} = (z_u, \ldots, z_t)$, $(s,z)_{u:t} = \{(s_r, z_r)\}_{r=u:t}$, $RS_{u:t} = (RS_u, \ldots, RS_t)$, $RF_{u:t} = (RF_u, \ldots, RF_t)$ whenever $u < t$, $\theta_\pi = (\{\pi_{ij,kl}\}_{i,j,k,l=1,\ldots,M})$, $\theta_u^{RS} = (\mu_1, \ldots, \mu_M, \nu_{11}, \ldots, \nu_{MM})$, $\theta_a^{RF} = (a_1, \ldots, a_M)$, $\theta_\sigma = (\gamma_1, \ldots, \gamma_M, \alpha_1, \ldots, \alpha_M, \beta_1, \ldots, \beta_M)$, $\theta_\tau = (\kappa_1, \ldots, \kappa_M, \omega_1, \ldots, \omega_M\ , \psi_1, \ldots, \psi_M)$ 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 1. *Suppose θ is known and assume that the observations are generated by the process described in (7). Then the conditional minimum variance hedge ratio at time t, is the solution to*

$$
h_t = \underset{h \in H}{\arg\min}\, Var(RS_t - hRF_t | \mathcal{F}_{t-1}^{\Theta}) \tag{9}
$$

which is given by

$$
h_t = \underbrace{\frac{Cov(\mu(s_t), a(z_t) | \mathcal{F}_{t-1}^{\Theta})}{Var(RF_t | \mathcal{F}_{t-1}^{\Theta})}}_{\textit{level-shift hedging}} + \underbrace{\sum_{i,j=1}^{M} \nu_{ij} w_{ij}}_{\textit{volatility hedging}}\ , \tag{10}
$$

where

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

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

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

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

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

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 AppendixA.

Proposition 1 states that the optimal hedge ratio at any point in time can be determined by two components. The first one, which we termed "level-shift hedging", 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 futures returns go on average in the same direction within the same regime then the hedge ratio increases. The second component, is customarily called "volatility hedging", 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. The role of the derivative's volatility in the hedging strategy is clear when the spot return level is regime independent. See Remark 2.

Remark 2. If $a(z_t)$ is constant, then the optimal hedge ratio in (10) reduces to the volatility hedging component

$$h_t = \sum_{i,j=1}^{M} \nu_{ij} \left(\frac{\left(\sum_{(s,z)_{1:t-1}} \tau_t^2(j) p(s_t = i, z_t = j|s_{t-1}, z_{t-1}, \theta) p((s,z)_{1:t-1}|\mathcal{F}_{t-1},\theta)\right)}{\sum_{i,j=1}^{M} \left(\sum_{(s,z)_{1:t-1}} \tau_t^2(j) p(s_t = i, z_t = j|s_{t-1}, z_{t-1}, \theta) p((s,z)_{1:t-1}|\mathcal{F}_{t-1},\theta)\right)} \right). \quad (11)$$

The effect of the dependence between hedging regimes s_t and futures variance states z_t on the hedge ratio is discussed in the following remarks. We consider the two cases of maximal dependence, i.e. $s_t = z_t$ (Remark 3) and maximal independence, both lagged and simultaneous independence (Remark 4).

Remark 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

$$h_t = \frac{Cov(\mu(s_t), a(s_t)|\mathcal{F}^{\Theta}_{t-1})}{Var(RF_t|\mathcal{F}^{\Theta}_{t-1})} + \sum_{j=1}^{M} \nu_j w_j, \quad (12)$$

where

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

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

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

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

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. If s_t and 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 (see Billio and Di Sanzo (2006)) given s_{t-1} (z_{t-1}), then, under the further assumption $\nu(s_t,z_t)=\nu(s_t)$, the result by Alizadeh and Nomikos (2004) is obtained, that is

$$\begin{aligned} h_t &= \sum_{j=1}^{M}\nu_j\left(\frac{\left(\sum_{s_{1:t-1}}p(s_t=j|s_{t-1},\theta)p(s_{1:t-1}|\mathcal{F}_{t-1},\theta)\right)}{\sum_{j=1}^{M}\left(\sum_{s_{1:t-1}}p(s_t=j|s_{t-1},\theta)p(s_{1:t-1}|\mathcal{F}_{t-1},\theta)\right)}\right) \\ &= \sum_{j=1}^{M}\nu_j p(s_t=j|\mathcal{F}_{1:t-1},\theta). \end{aligned} \tag{13}$$

We expect a more flexible and efficient hedging strategy 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 2, the model parameters in Equation (10) are not known in practice. In this respect, a natural approach to solving this problem will be to apply the PCE principle. Alternatively, following the Bayesian paradigm outlined above we have the following proposition.

Proposition 5. *Assume that the observations are generated by the process described in (7). Then under certain regularity conditions the Bayesian conditional minimum hedge ratio at time t is the solution to*

$$h_t^{BAY} = \arg\min_{h\in H}\{E(Var(RS_t-hRF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1}))\} \tag{14}$$

which is given by

$$h_t^{BAY} = \underbrace{\frac{\int_\Theta[Cov(\mu(s_t),a(z_t)|\mathcal{F}^{\Theta}_{t-1})]p(\theta|\mathbf{y}_{1:t-1})d\theta}{\int_\Theta[Var(RF_t|\mathcal{F}^{\Theta}_{t-1})]p(\theta|\mathbf{y}_{1:t-1})d\theta}}_{\text{integrated level-shift hedging}} + \underbrace{\sum_{i,j=1}^{M}\int_\Theta \nu_{ij}w_{ij}(\theta|\mathbf{y}_{1:t-1})d\theta}_{\text{integrated volatility hedging}}, \tag{15}$$

where,

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

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

$\tau_t^2(k)=\kappa_k+\omega_k\xi_{t-1}^2+\phi_k\tau_{t-1}^2$ *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 AppendixB.

Similar to Proposition (1), Proposition (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.

2.2. Computational issues

An important ingredient needed in the computation of the optimal hedge ratio in (15) is the posterior distribution of the augmented parameter vector $p((s,z)_{1:t-1},\theta|\mathbf{y}_{1:t-1})$, $\quad t=\bar{t},\dots,T$. These quantities cannot be identified with any known distribution. This limitation makes the evaluation of (15) non-trivial. 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 consists of approximating the posterior distribution of the unknown parameters vector given past observations, while the second part involves evaluating the hedge ratio.

Following Billio et al. (2014), we propose 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 proposed 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 samples generated by this MCMC algorithm are used in the second stage for approximating the moments in (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 the identifiability restriction: $\gamma_1<\gamma_2<\cdots<\gamma_M$, $\kappa_1<\kappa_2<\cdots<\kappa_M$. In order to generate samples from the posterior density of the augmented parameter vectors:

$$f(\theta,(s,z)_{1:t}|RS_{1:t},RF_{1:t})\propto f(RS_{1:t}|(s,z)_{1:t},\theta,RF_{1:t})f(RF_{1:t}|(s,z)_{1:t},\theta)p((s,z)_{1:t}|\theta)f(\theta) \tag{16}$$

our Gibbs sampler iterates over the following steps:

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

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

$$p((s,z)_{1:t}|\theta,RS_{1:t},RF_{1:t})\propto f(RS_{1:t}|RF_{1:t},\theta,(s,z)_{1:t})f(RF_{1:t}|\theta,(s,z)_{1:t}) \tag{17}$$

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. (2014) for alternative approximations. Samples of the state trajectory are then drawn by Forward Filter Backward sampling scheme. Details of the proposal construction and the MH algorithm are given in Appendix AppendixC.

In the second stage, G MCMC samples from $p((s,z)_{1:t-1}|\theta,\mathcal{F}_{t-1}^{\Theta})$ are used to approximate the hedge ratio:

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

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

It is worth noting that the decision problem characterized by Proposition 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 an 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 remains feasible in practice since the computation of hedging ratio only requires about an hour on daily basis. Alternative procedures such as the sequential MCMC or sequential Monte Carlo may be used to reduce the computing time when a timely updating of the hedge ratio is required.

3. An application to energy markets

The goal of this section is twofold. First, we aim to apply our model to provide empirical evidence of the effects of the recent financial crisis on the crude oil markets. Second we want to asses the efficiency of the proposed hedging models and to compare them.

3.1. Hedging on the crude oil market

We consider 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 (1) displays the sample path of the crude oil squared returns on spot and futures. We observe volatility clustering, which calls for the use of MS-GARCH models.

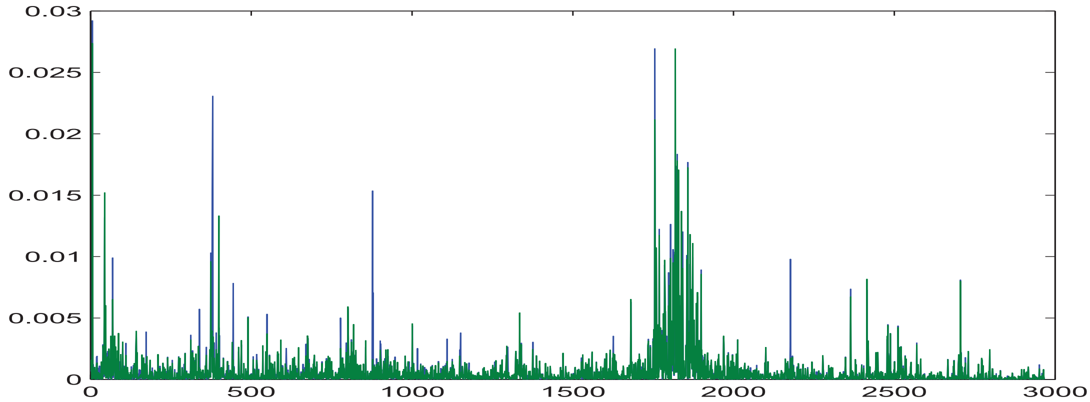


Figure 1: Graphs for daily squared returns on WTI crude oil spot and futures from September 14/09/2001 to 31/07/2013

Before proceeding with the evaluation of 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 (i.e. independent, dependent and same state variable). This analysis will allow us to investigate some volatility features of the spot and futures oil markets. Moreover, the specification used for the bivariate spot and futures MS-GARCH model will allow us to draw some conclusions on the optimal hedging strategy to use in these markets.

We perform 10000 Gibbs iterations after convergence according to the Geweke's diagnostic (Geweke (1992)). To reduce serial correlation of the draws, we consider every 10th draw after convergence of the Gibbs iteration to obtain the results presented below. Table 1 to 4 show the estimation results for two-state models using the full sample of observation described above. From the estimated parameters in Table 1, regime 1 may be labeled as the low volatility state. With the exception of the constrained multichain MS-GARCH model, we observe from Table 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 a low volatility regime. This result 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 2 show that the probability that the hedged portfolio and futures return simultaneously remain in the high regime is very low. Whereas, from Table 4 we observe that the single chain MS-GARCH model gives a relatively high probability for the two variables being in the high state simultaneously. The implication of this observation is that, when possible misalignments 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 of the volatility. Nevertheless, a robust 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 likely that this scenario will be maintained in the next period. Also, Table 2 suggests that when the returns on the hedged portfolio are 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 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 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

Given the current volatility states across markets, the unconstrained multichain MS-GARCH framework provides an avenue to studying the influence of a change in the state of the futures on the probability of the

Table 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: 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

hedged portfolio remaining in the same regime, and vice versa, 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 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 a hedged portfolio staying in regime 1, when the futures is in regime 1 in the previous month, is 0.92, but decreases to 0.72 when the futures is in regime 2. In a similar way, the futures remains in regime 1 with a 98% chance when the hedged portfolio is in the same regime, but switches to regime 2 with a probability equal to 31% when the hedged portfolio is in regime 2.

Table 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

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

Differently to the single chain MS-GARCH model with two correlations, the multichain MS-GARCH models have four correlation regimes at each point in time. To get 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 variances. 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 brief, we find that the correlation of spot and futures return series tends 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 equal, respectively, to 0.979 and 0.822. These values are somewhere between the highest and the lowest correlation estimated from the multichain MS-GARCH model. Overall, more model flexibility may be achieved with the unconstrained MS-GARCH model, since it is the one with the widest correlation range, among the models considered in this study.

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.

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. Moreover, the sequential analysis will allow us to provide evidence of time-changes in the volatility transmission mechanisms, and in the correlation between the two markets. We will exploit the hedge ratio interpretation of one of the parameters used in our model to study the effects on the optimal hedging strategy due to possible time-variations in the volatility and correlation structures of these markets. For each hedging model, an out-of-sample analysis of its hedging performance with daily re-balancing is carried out. On a 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 correspond, respectively, to the period before, during, and after the 2008/2009 global financial crisis.

Figure 2 - 6 shows how the different regime-specific hedge ratios, conditional variances 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 2(a)), and high variability in the conditional variance of the futures (see 3(a)), prior to the 2008/2009 global financial crisis. However, after the 2008/2009 global financial crisis, we observe a clear separation of the hedge ratios (2(e)) into two groups determining the change in the hidden process on the futures returns. Also, the variability of the conditional variance of the futures is very low (2(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 their experience during financial crisis. The application of the single chain MS-GARCH model tends to suffer from an under-or over-estimation problem arising from the use of a single chain (see 4(a)-4(e)). In the case of the constrained MS-GARCH model, there is 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 model is more flexible as it produces a wider range of values for the hedge ratio. Overall, the unconstrained MS-GARCH model seems to perform best among the set of models under comparison. Also, the high probability of staying in the low volatility regime implies low transaction costs because the investor only needs to re-balance his/her portfolio occasionally.

In the above exercise, it is assumed that the prevailing state of the world is known. However the current state of the world cannot be correctly identified by the hedger. In this situation, the mean hedge ratio implemented is:

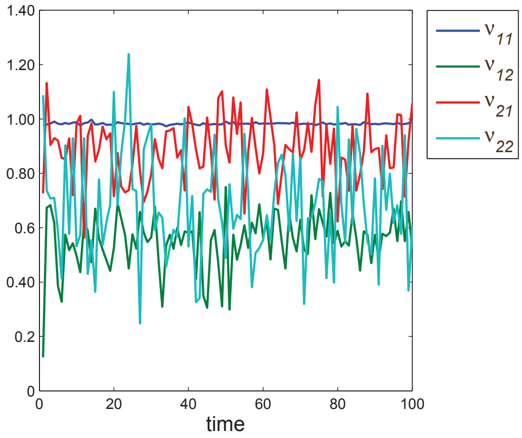
$$h_t = E_\theta E[\nu(s_t, z_t)|\mathcal{F}_{t-1}, \Theta] = E_\theta \left[\sum_{m,m'=1}^{M} \nu_m p(s_t = m, z_t = m'|\mathcal{F}_{t-1}, \Theta) \right], \tag{18}$$

and has been approximated using,

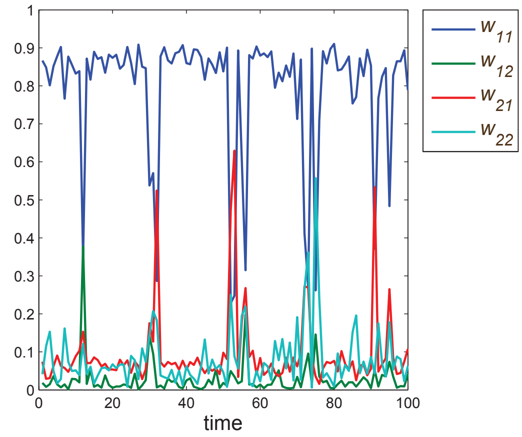
$$h_t = \frac{1}{G} \sum_{i=1}^{G} \sum_{m,n=1}^{M} \nu_{m,n}^{(i)} p^{(i)}(s_t = m, z_t = n|\mathcal{F}_{t-1}, \Theta), \tag{19}$$

where G is the number of Gibbs samples.

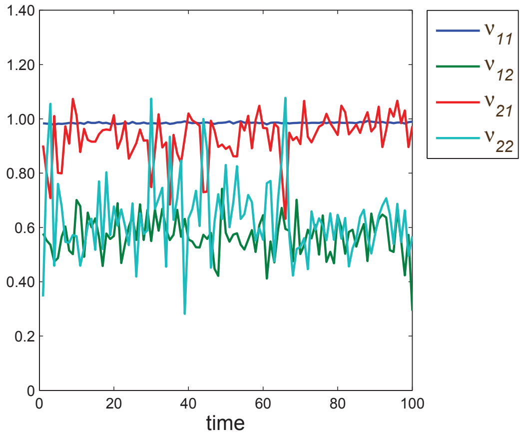
In Figure 7, we report the estimation results for each model and compare them with the OLS hedge ratio over the three subsamples. The MS-GARCH hedge ratios display similar time-varying characteristics. However, we occasionally observe that the time varying hedge ratios fall below the OLS hedge ratio. Also, the hedge ratios are observed to shift closer to 1 after the global financial crisis. This confirms our earlier intuition that hedgers seem to have become more careful in their investment decisions after the global financial crisis.



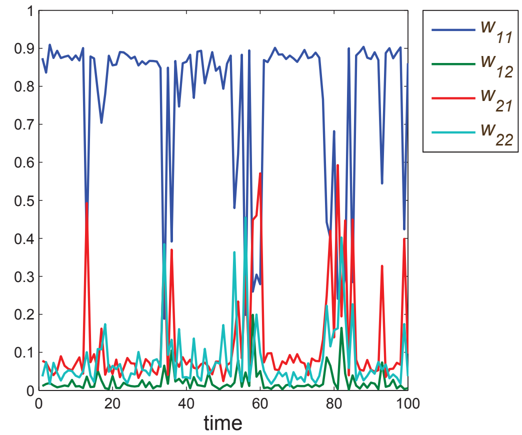
(a) Regime specific hedge ratio before the crisis



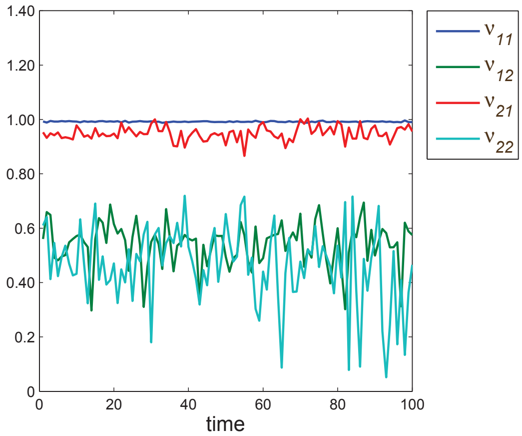
(b) Regime specific weights before the crisis



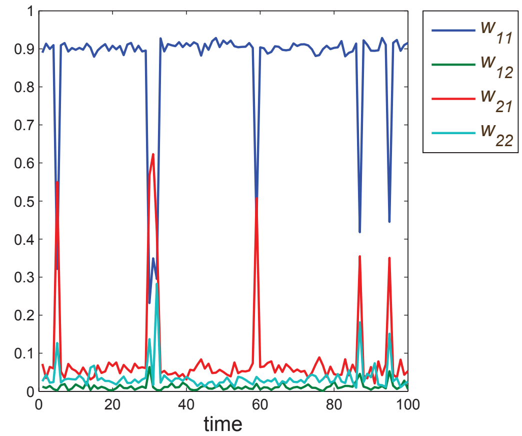
(c) Regime specific hedge ratio during the crisis



(d) Regime specific weights during the crisis

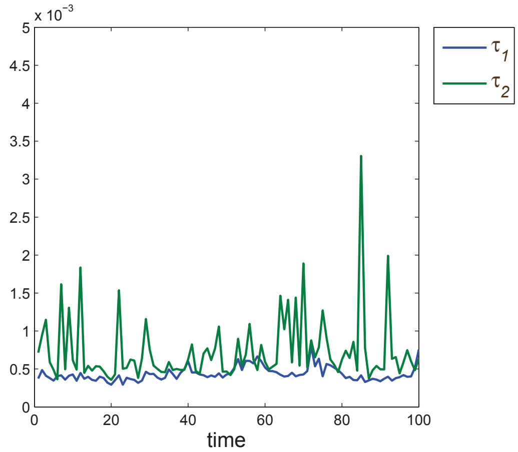


(e) Regime specific hedge ratio after the crisis

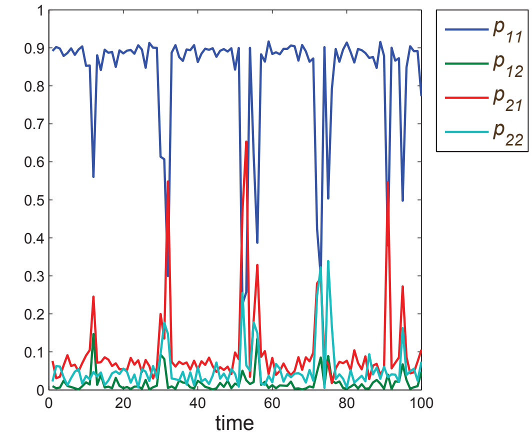


(f) Regime specific weights after the crisis

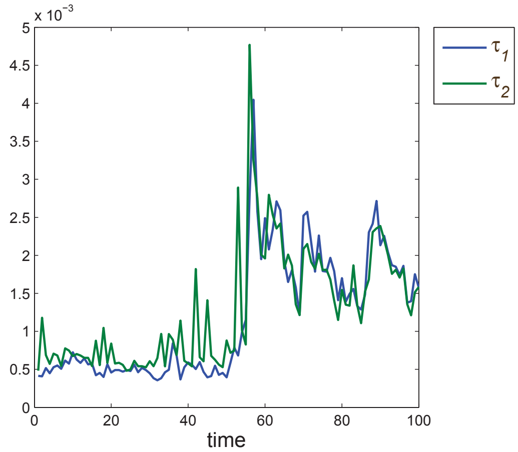
Figure 2: Regime specific hedge ratio and corresponding weights for the unconstrained multichain MS-GARCH model (MC-f-MSGARCH). first row 08/08/2006 to 03/01/2007; second row 01/10/2008 to 25/03/2009; third row and 15/02/2013 to 31/07/2013.



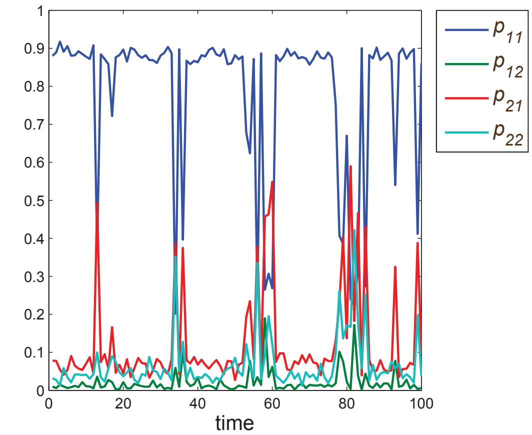
(a) Regime specific conditional variance before the crisis



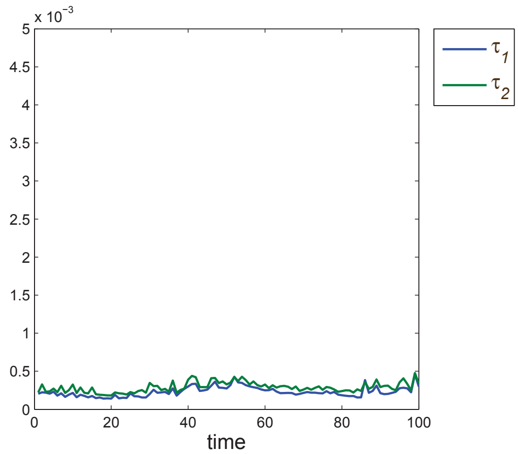
(b) Regime specific prediction probability before the crisis



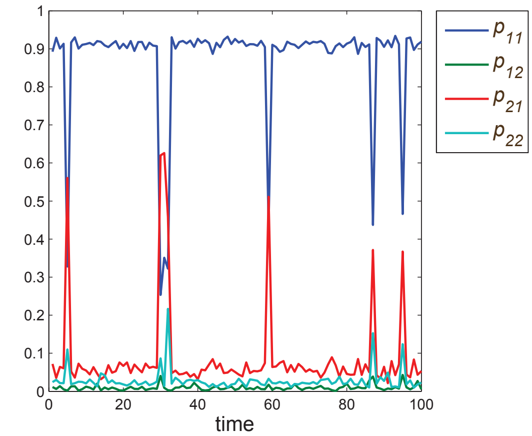
(c) Regime specific conditional variance during the crisis



(d) Regime specific prediction probability during the crisis

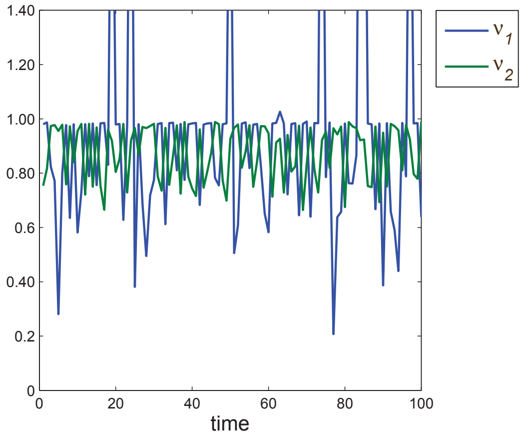


(e) Regime specific conditional variance after the crisis

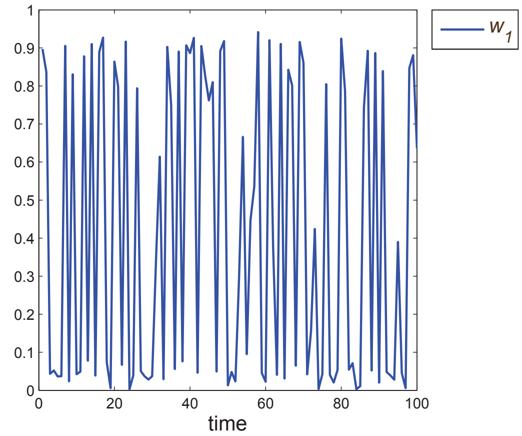


(f) Regime specific prediction probability after the crisis

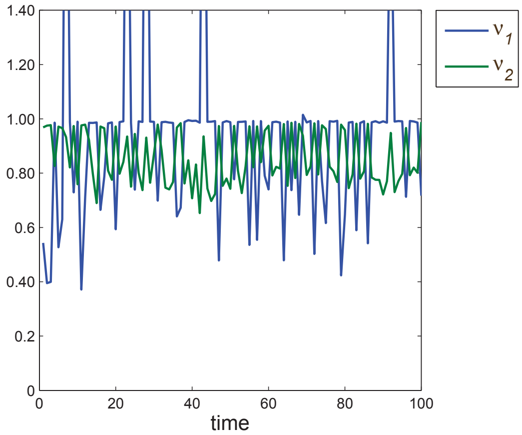
Figure 3: Regime specific conditional variance and corresponding predicted probabilities for the unconstrained multichain MS-GARCH model (MC-f-MSGARCH). first row 08/08/2006 to 03/01/2007; second row 01/10/2008 to 25/03/2009; third row and 15/02/2013 to 31/07/2013.



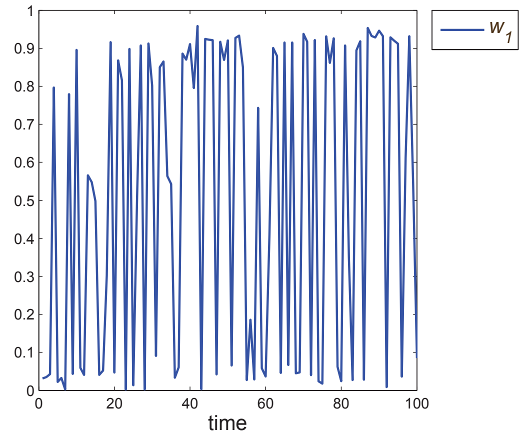
(a) Regime specific hedge ratio before the crisis



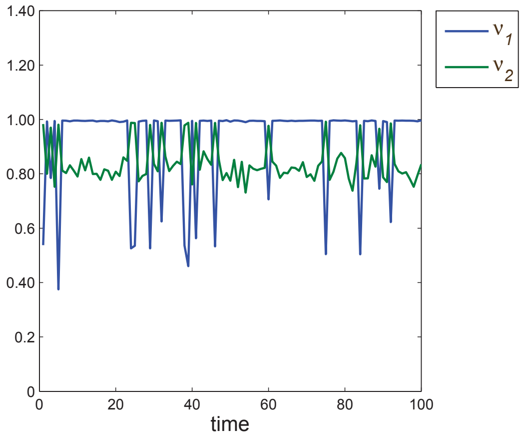
(b) Regime specific weights before the crisis



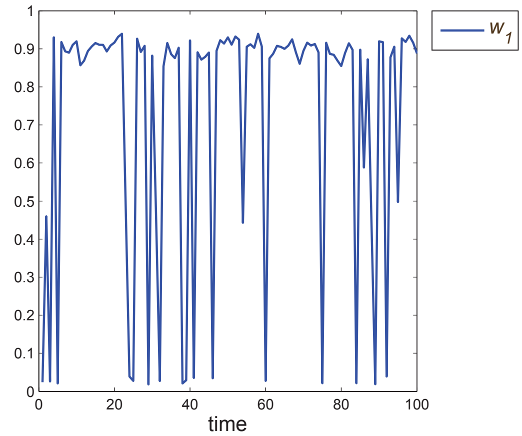
(c) Regime specific hedge ratio during the crisis



(d) Regime specific weights during the crisis

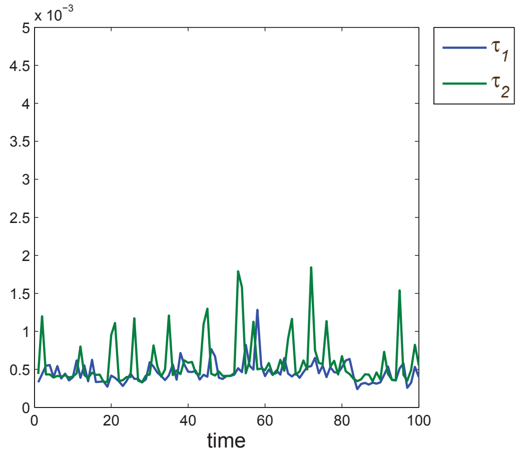


(e) Regime specific hedge ratio after the crisis

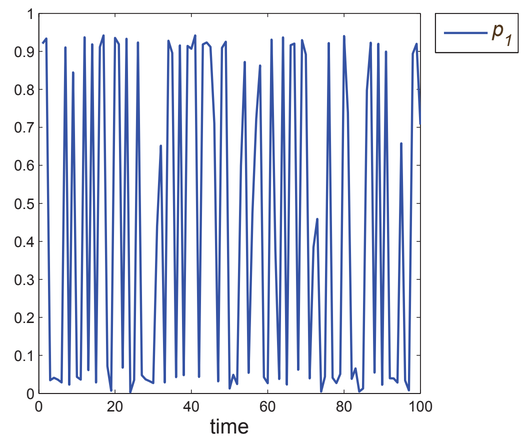


(f) Regime specific weights after the crisis

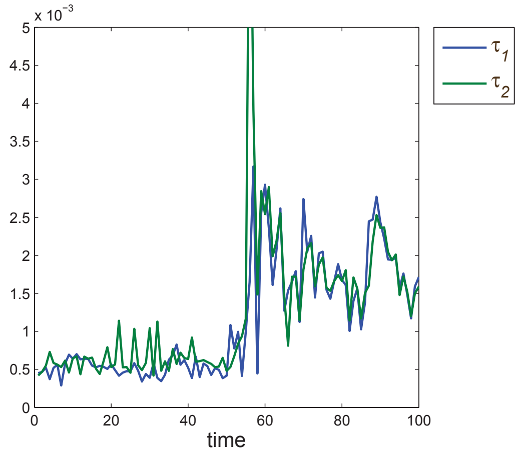
Figure 4: Regime specific hedge ratio and corresponding weights for the single chain MS-GARCH model (SC-MSGARCH). first row 08/08/2006 to 03/01/2007; second row 01/10/2008 to 25/03/2009; third row and 15/02/2013 to 31/07/2013.



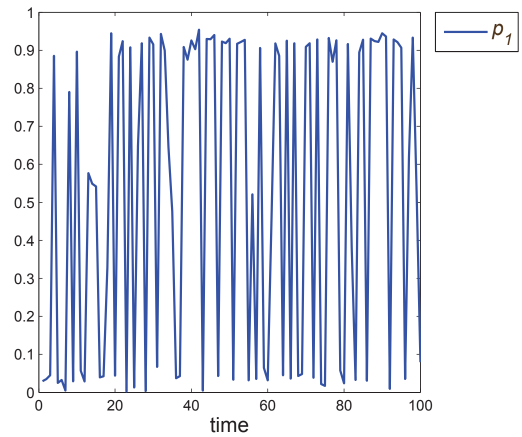
(a) Regime specific conditional variance before the crisis



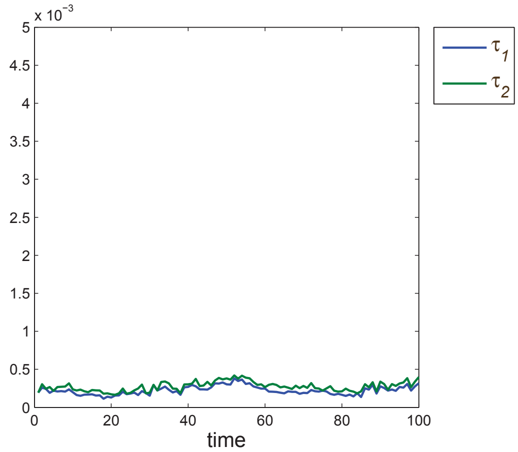
(b) Regime specific prediction probability before the crisis



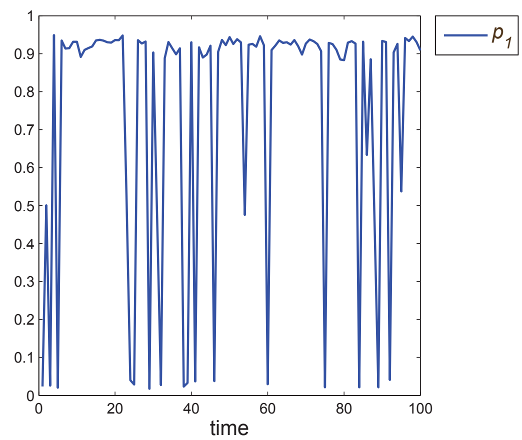
(c) Regime specific conditional variance during the crisis



(d) Regime specific prediction probability during the crisis

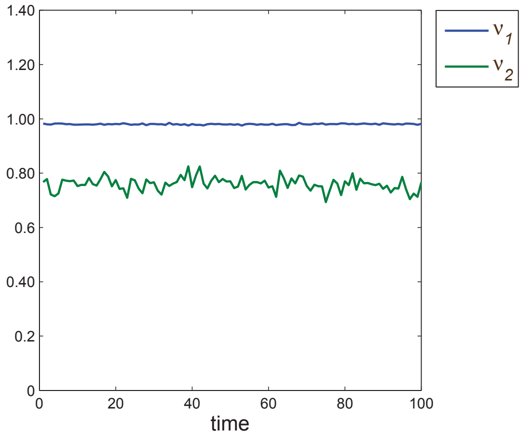


(e) Regime specific conditional variance after the crisis

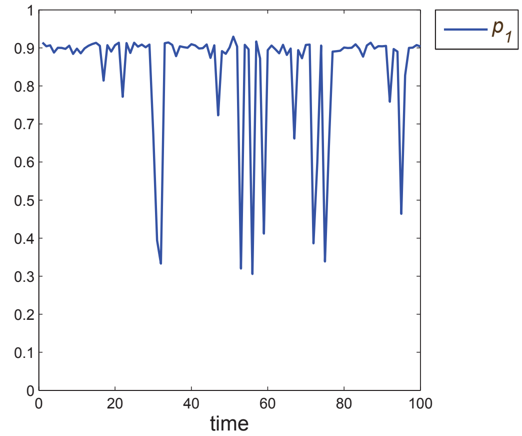


(f) Regime specific prediction probability after the crisis

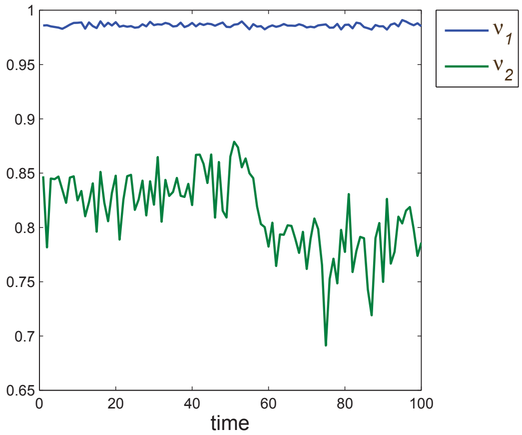
Figure 5: Regime specific conditional variance and the corresponding prediction probabilities for the single chain MS-GARCH model (SC-MSGARCH). first row 08/08/2006 to 03/01/2007; second row 01/10/2008 to 25/03/2009; third row and 15/02/2013 to 31/07/2013.



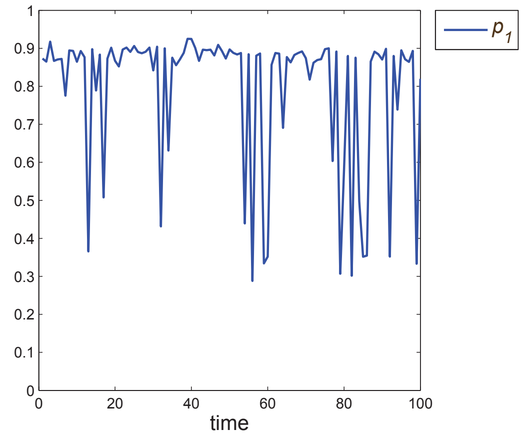
(a) Regime specific hedge ratio before the crisis



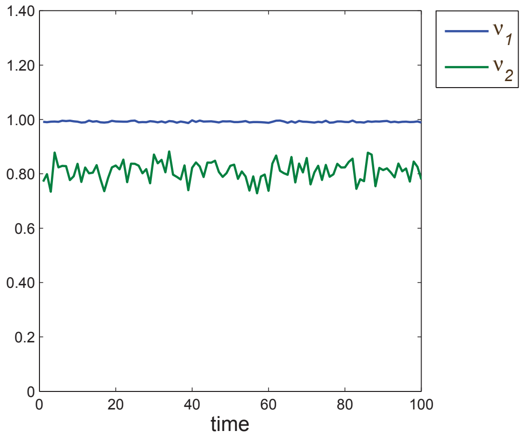
(b) Regime specific weights before the crisis



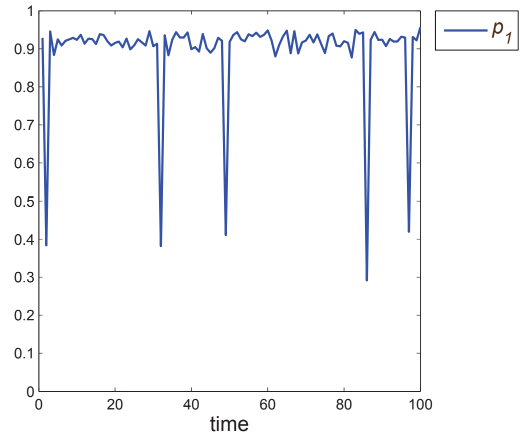
(c) Regime specific hedge ratio during the crisis



(d) Regime specific weights during the crisis

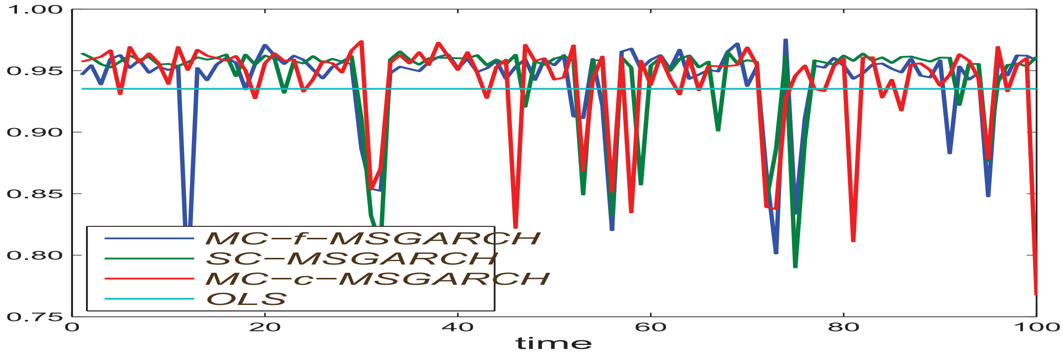


(e) Regime specific hedge ratio after the crisis

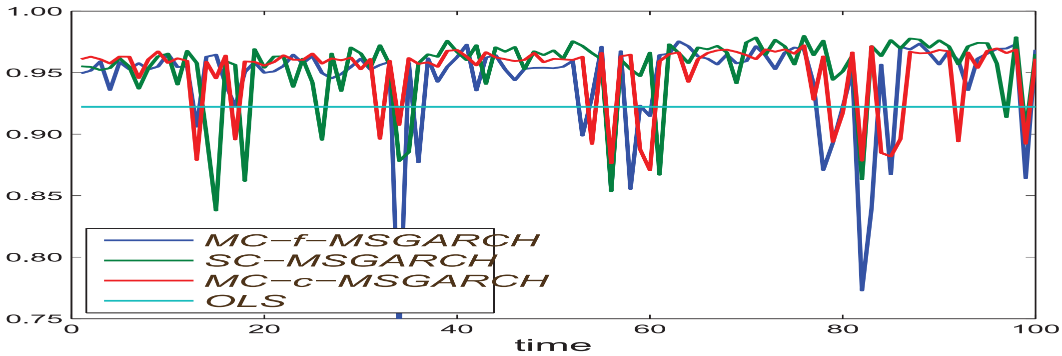


(f) Regime specific weights after the crisis

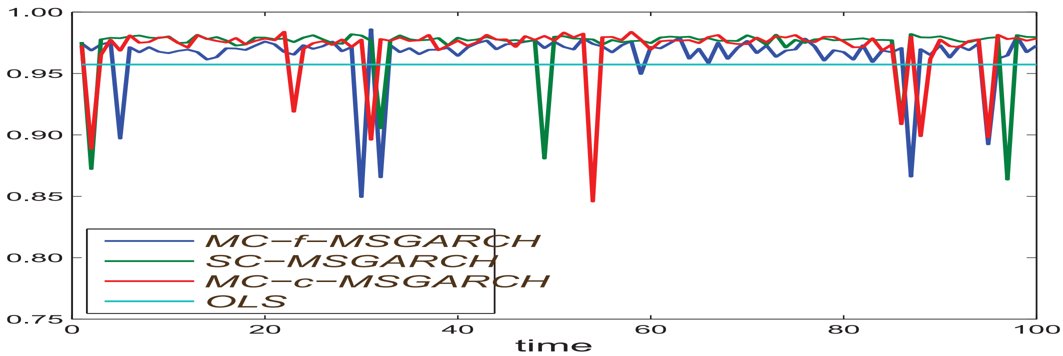
Figure 6: Regime specific hedge ratio and corresponding prediction probabilities for the constrained multi-chain MS-GARCH model (MC-c-MSGARCH). first row 08/08/2006 to 03/01/2007; second row 01/10/2008 to 25/03/2009; third row and 15/02/2013 to 31/07/2013.



(a) Hedge ratio before financial crisis



(b) Hedge ratio during the financial crisis



(c) Hedge ratio after the financial crisis

Figure 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.

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 subsample period. In mathematical forms, we evaluate

$$Var(RS_t - h_t^* RF_t) \tag{20}$$

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

$$\frac{Var(\text{OLS}) - Var(\text{MS-GARCH})}{Var(\text{OLS})} \times 100, \tag{21}$$

where Var(OLS) and Var(MS-GARCH) are respectively the variance of the returns on the hedged portfolio (Equation (20)) estimated using hedge ratios obtained from the OLS and MS-GARCH models. A positive value of 21 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 21 are considered. The first is the average hedge ratio given by 19, 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 6: Hedging Effectiveness of MS-GARCH against Constant Hedge ratio.

	$h_t^* = E[\nu(s_t, z_t)]$			$h_t^* = E[\nu_t \| \hat{s}_{t-1}, \hat{z}_{t-1}]$			$h_t^* = \nu(\hat{s}_t.\hat{z}_t)$		
	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 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, seems 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 21, 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 performs better than the unconstrained multichain model. This is in contrast with our observation when the average hedge ratio is applied. This result 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 the observation of Sephton (1998) who finds that the Regime Switching strategy outperforms both OLS and GARCH strategies in the low variance state, but performs far worse than either strategy in the high variance state. This is an indication that multichain Markov-switching models have the potentials to compete favourably 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 insights into the relative advantages of the multichain regime switching model over other models.

4. Conclusion

In this paper we propose a new Bayesian multichain MS-GARCH model with dependent chains. We apply the model to hedging in energy markets, thus extending the existing literature on MV hedging. The proposed parameterization of the multichain MS-GARCH model allows for a straightforward interpretation of the parameters of the models as level-shift and variance-covariance hedging components. Both the Bayesian model and the approach to inference allow us to easily account for parameter uncertainty in the hedging decision. We apply this multichain MS-GARCH to estimate a state-dependent time-varying minimum variance hedge ratio and investigate the effect of relaxing the assumption of common switching dynamic on the effectiveness of the hedging strategy. The practical implementation of our hedging model to crude oil spot and futures markets shows strong evidence in favour of the unconstrained multichain MS-GARCH model when compared to other models, in terms of hedging effectiveness. Nevertheless, a sequential model comparison on the three sub-periods, i.e. before, during and after the 2008/2009 global financial crisis, provides evidence in support of the best MS-GARCH models, before and during the financial crisis. In the period after the crisis, the reduction in the volatility level makes the MS-GARCH less appealing than the standard OLS approach. This paper offers many opportunities for further research.

First, the hedging strategy and measures of hedging effectiveness as considered in this paper ignores transaction cost. Generalizing our hedging framework by incorporating transaction cost may provide better insight into the practical usefulness of the proposed strategy. Also, the performance of our hedging strategy could be further enriched by accounting for model uncertainty. This may be achieved by embedding Bayesian Model Averaging (BMA) into our hedging strategy. Lastly, in order to create a balance between variance reduction and incremental transaction cost, alternative measures of hedging effectiveness such as a utility framework should be considered. See Alizadeh and Nomikos (2004) for illustration.

AppendixA. Proof to Proposition 1:

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

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 the on RS_t and RF_t. Therefore, our problem reduces to

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

From the first order condition, we have

$$\begin{aligned} h_t &= \frac{Cov(RS_t, RF_t|\mathcal{F}^{\Theta}_{t-1})}{Var(RF_t|\mathcal{F}^{\Theta}_{t-1})}, \\ &= \frac{Cov(\mu(s_t), RF_t|\mathcal{F}^{\Theta}_{t-1})}{V(RF_t|\mathcal{F}^{\Theta}_{t-1})} + \frac{Cov(\nu(s_t,z_t)RF_t, RF_t|\mathcal{F}^{\Theta}_{t-1})}{V(RF_t|\mathcal{F}^{\Theta}_{t-1})}. \end{aligned} \tag{A.3}$$

where,

$$\begin{aligned} Cov(\nu(s_t,z_t)RF_t, RF_t|\mathcal{F}^{\Theta}_{t-1}) &= E[\nu(s_t,z_t)RF_t^2|\mathcal{F}^{\Theta}_{t-1}] - E[\nu(s_t,z_t)RF_t|\mathcal{F}^{\Theta}_{t-1}]E[RF_t|\mathcal{F}^{\Theta}_{t-1}] \\ &= E[\nu(s_t,z_t)(a(z_t)+\tau_t\zeta_t)^2|\mathcal{F}^{\Theta}_{t-1}] \\ &\quad - E[\nu(s_t,z_t)(a(z_t)+\tau_t\zeta_t)|\mathcal{F}^{\Theta}_{t-1}]E[(a(z_t)+\tau_t\zeta_t)|\mathcal{F}^{\Theta}_{t-1}], \\ &\overset{iid\ \zeta_t}{=} E[\nu(s_t,z_t)(a(z_t)^2+\tau_t^2)|\mathcal{F}^{\Theta}_{t-1}] - E[\nu(s_t,z_t)a(z_t)|\mathcal{F}^{\Theta}_{t-1}]E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}], \\ &= E[\nu(s_t,z_t)(a(z_t)^2+\tau_t^2 - a(z_t)E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])|\mathcal{F}^{\Theta}_{t-1}]. \end{aligned} \tag{A.4}$$

Then by law of iterated expectation, we have

$$\begin{aligned} &Cov(\nu(s_t,z_t)RF_t, RF_t|\mathcal{F}^{\Theta}_{t-1}) \\ &= E\left(\sum_{i,j=1}^{M} \nu_{ij}(a_j^2+\tau_t^2(j) - a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])p(s_t=i, z_t=j|(s,z)_{1:t-1}, \mathcal{F}_{t-1}, \theta)|\mathcal{F}^{\Theta}_{t-1}\right) \\ &= E\left(\sum_{i,j=1}^{M} \nu_{ij}(a_j^2+\tau_t^2(j) - a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}|\mathcal{F}^{\Theta}_{t-1}\right) \\ &= \sum_{(s,z)_{1:t-1}} \left(\sum_{i,j=1}^{M} \nu_{ij}(a_j^2+\tau_t^2(j) - a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}\right) p((s,z)_{1:t-1}|\mathcal{F}_{t-1},\theta) \\ &= \sum_{i,j=1}^{M} \nu_{ij} \left(\sum_{(s,z)_{1:t-1}} (a_j^2+\tau_t^2(j) - a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1},\theta)\right), \end{aligned} \tag{A.5}$$

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,

$$V(RF_t|\mathcal{F}^{\Theta}_{t-1}) = \sum_{i,j=1}^{M} \left(\sum_{(s,z)_{1:t-1}} (a_j^2+\tau_t^2(j) - a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1},\theta)\right), \tag{A.6}$$

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

$$Cov(\mu(s_t), RF_t|\mathcal{F}^{\Theta}_{t-1}) = Cov(\mu(s_t), a(z_t)|\mathcal{F}^{\Theta}_{t-1}) =$$

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

$\tau_t^2(j) = \kappa_j + \omega_j \xi_{t-1}^2 + \phi_j \tau_{t-1}^2$ for $j = 1, \dots, M$ and t=1,..., T. The result follows immediately by substituting these quantities into (A.3) and letting

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

AppendixB. Proof to Proposition 5:

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

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

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

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

$$\begin{aligned} h_t &= \frac{E(Cov(RS_t, RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}{E(Var(RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}, \\ &= \frac{E(Cov(\mu(s_t), RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}{E(V(RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})} + \frac{E(Cov(\nu(s_t,z_t)RF_t, RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}{E(V(RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})}. \end{aligned} \tag{B.3}$$

where,

$$\begin{aligned} &E(Cov(\nu(s_t,z_t)RF_t, RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1}) \\ &= \int_{\Theta} \left(\sum_{i,j=1}^{M} \nu_{ij} \left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..} p((s,z)_{1:t-1}|\mathcal{F}_{t-1},\theta)\right)\right) p(\theta|\mathbf{y}_{1:t-1})d\theta \\ &= \sum_{i,j=1}^{M} \left(\int_{\Theta} \nu_{ij} \left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..} p((s,z)_{1:t-1}|\mathcal{F}_{t-1},\theta)\right) p(\theta|\mathbf{y}_{1:t-1})d\theta\right) \\ &= \sum_{i,j=1}^{M} \left(\int_{\Theta} \nu_{ij} \left(\sum_{(s,z)_{1:t-1}} (a_j^2 + \tau_t^2(j) - a_j E[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..} p((s,z)_{1:t-1},\theta,|\mathcal{F}_{t-1},\theta)\right) d\theta\right), \end{aligned} \tag{B.4}$$

and

$$
\begin{aligned}
&E(V(RF_t|\mathcal{F}^{\Theta}_{t-1})|\mathcal{F}_{t-1})\\
&=\int_{\Theta}\left(\sum_{i,j=1}^{M}\left(\sum_{(s,z)_{1:t-1}}(a_j^2+\tau_t^2(j)-a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1},\theta)\right)\right)p(\theta|\mathbf{y}_{1:t-1})d\theta\\
&=\sum_{i,j=1}^{M}\left(\int_{\Theta}\left(\sum_{(s,z)_{1:t-1}}(a_j^2+\tau_t^2(j)-a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1}|\mathcal{F}_{t-1},\theta)\right)p(\theta|\mathbf{y}_{1:t-1})d\theta\right) \qquad \text{(B.5)}\\
&=\sum_{i,j=1}^{M}\left(\int_{\Theta}\left(\sum_{(s,z)_{1:t-1}}(a_j^2+\tau_t^2(j)-a_jE[a(z_t)|\mathcal{F}^{\Theta}_{t-1}])\pi_{ij,..}p((s,z)_{1:t-1},\theta|\mathcal{F}_{t-1},\theta)\right)d\theta\right).
\end{aligned}
$$

$\tau_t^2(j)=\kappa_j+\omega_j\xi_{t-1}^2+\phi_j\tau_{t-1}^2$ for $j=1,\ldots,M$ and t=1,..., T. The result follows immediately by substituting these quantities into (B.3) with

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

AppendixC. Computational details

AppendixC.1. Forward filtering

Let the proposal distribution be denoted by

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

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

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

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

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

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

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

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

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

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

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

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

AppendixC.2. 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

$$f(\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau | s_{1:t}^{(r)}, \pi, y_{1:t}) \propto f(\theta_u^{RS}, \theta_a^{RF}, \theta_\sigma, \theta_\tau)$$
$$\times \prod_{i=1}^{t} \frac{1}{\sigma_i} \exp\left(-\frac{(RS_i - \mu(s_i) - \nu(s_i) RF_i)^2}{2\sigma_i^2} \right)$$
$$\times \prod_{i=1}^{i} \frac{1}{\tau_i} \exp\left(\frac{-(RF_i - a(s_i))^2}{2\tau_i^2} \right) \tag{C.3}$$

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 $\xi_i = (\mathbb{I}_{s_i=1}, \ldots, \mathbb{I}_{s_i=M})'$ 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.

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

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

$$\begin{aligned} & f(\theta_\sigma|\xi_{1:T}^{(r)}, \mu^{(r)}, y_{1:T}) \propto \\ & \propto \frac{1}{|\mathbf{V}|^{\frac{1}{2}}}\exp\left(-\frac{\mathbf{w}'\mathbf{V}^{-1}\mathbf{w}}{2}\right)\mathbb{I}_{\{\gamma_1>0\ldots,\gamma_M>0,0<\alpha_1<1,\ldots,0<\alpha_M<1,0<\beta_1<1\ldots,0<\beta_M<1\}} \\ & \propto \mathcal{N}_{3M}(m_\sigma, \Sigma_\sigma)\mathbb{I}_{\{\gamma_1>0\ldots,\gamma_M>0,0<\alpha_1<1,\ldots,0<\alpha_M<1,0<\beta_1<1\ldots,0<\beta_M<1\}}, \end{aligned} \quad \text{(C.5)}$$

where

$$\begin{aligned} \Sigma_\sigma &= (\nabla'\mathbf{V}^{-1}\nabla)^{-1} \\ m_\sigma &= \Sigma\nabla'\mathbf{V}^{-1}\mathbf{r}^*. \end{aligned} \quad \text{(C.6)}$$

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

AppendixC.3. Pseudo code of the hedging algorithm

Algorithm 1: Posterior approximation

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

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,

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

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

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

   where

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

Algorithm 2: Hedging

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

1. Compute the moments and substitute into 2.2

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

Acknowledgments

We thank seminar participants at the SIS 2013 Statistical Conference, Brescia, 2013. Authors' research is supported by funding from the European Union, Seventh Framework Programme FP7/2007-2013 under grant agreement SYRTO-SSH-2012-320270, and by the Italian Ministry of Education, University and Research (MIUR) PRIN 2010-11 grant MISURA.

References

Alizadeh AH, Nomikos NK. A Markov regime switching approach for hedging stock indices. Journal of Futures Markets 2004;24:649 –74.

Alizadeh AH, Nomikos NK, Pouliasis PK. A Markov regime switching approach for hedging energy commodities. Journal of Banking & Finance 2008;32:1970 –83.

Amzal B, Bois FY, Parent E, Robert CP. Bayesian-optimal design via interacting particle systems. Journal of the American Statistical Association 2006;101(474):773–85.

Ardia D. 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.

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

Billio M, Casarin R, Osuntuyi A. Efficient Gibbs sampling for Markov switching GARCH models. Computational Statistics and Data Analysis, forthcoming 2014;.

Billio M, Di Sanzo S. Granger-causality in Markov switching models. Working Paper 20/06 University of Ca' Foscari 2006;.

Chang CY, Lai JY, Chuang IY. Futures hedging effectiveness under the segmentation of bear/bull energy markets. Energy Economics 2010;32:442 –9.

Chen S, Lee C, Shrestha K. Futures hedge ratios: a review. The Quarterly Review of Economics and Finance 2003;43:433 –65.

DeGroot MH. Optimal statistical decisions. volume 82. Wiley-Interscience, 2005.

Ederington L. The hedging performance of the new futures markets. Journal of Finance 1979;34(1):157 –70.

Geweke J. Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. Bayesian Statistics 4. Oxford: Oxford University Press; 1992. p. 169–93.

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

Guidolin M, Rinaldi F. Ambiguity in asset pricing and portfolio choice: a review of the literature. Theory and Decision 2013;74:183 – 217.

Haigh MS, Holt MT. Crack spread hedging: accounting for time-varying volatility spillovers in the energy futures markets. Journal of Applied Econometrics 2002;17:269 –89.

Johnson LL. The theory of hedging and speculation in commodity futures. Review of Economic Studies 1960;27(3):139 –51.

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

Lee H, Yoder JK. A bivariate Markov regime switching GARCH approach to estimate time varying minimum variance hedge ratio. Applied Economics 2007a;39(10):1253 –65.

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

Lee HT, Yoder JK, Mittelhammer RC, McCluskey JJ. A random coefficient autoregressive Markov regime switching model for dynamic futures hedging. The Journal of Futures Markets 2006;26(2):103 –29.

Lence SH, Hayes DJ. The empirical minimum-variance hedge. American Journal of Agricultural Economics 1994a;76(1):94 – 104.

Lence SH, Hayes DJ. Parameter-based decision marketing under estimation risk: An application to futures trading. Journal of Finance 1994b;49(1):345 –57.

Lien D. A note on the superiority of the ols hedge ratio. Journal of Futures Markets 2005;25:1121 –6.

Lien D. A note on the hedging effectiveness of GARCH models. International Reviews of Economics and Finance 2009;18:110 –2.

Müller P. Simulation-based optimal design. Bayesian statistics 1999;6:459–74.

Müller P, Sansó B, Iorio MD. Optimal Bayesian design by inhomogeneous Markov chain simulation. Journal of the American Statistical Association 2004;99(467):788–98.

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

Nakatsuma T. A Markov-chain sampling algorithm for GARCH models. Studies in Nonlinear Dynamics and Econometrics 1998;3(2):107–17.

Raiffa H, Schlaifer R. Applied Statistical Decision Theory. Harvard University Press, 1961.

Sephton P. GARCH and Markov hedging at the Winnipeg Commodity Exchange. Canadian Journal of Agricultural Economics 1998;46:117 – 26.

Sheu HJ, Lee HT. Optimal futures hedging under multichain Markov regime switching. Journal of Futures Markets 2012;.