

Forecasting Time-Varying Covariance with a Robust Bayesian Threshold Model

CHIH-CHIANG WU^{1*} AND JACK C. LEE²

¹ *Department of Finance, Yuan Ze University, Taoyuan, Taiwan*

² *Graduate Institute of Finance, National Chiao Tung University, Hsinchu, Taiwan*

ABSTRACT

This paper proposes a robust multivariate threshold vector autoregressive model with generalized autoregressive conditional heteroskedasticities and dynamic conditional correlations to describe conditional mean, volatility and correlation asymmetries in financial markets. In addition, the threshold variable for regime switching is formulated as a weighted average of endogenous variables to eliminate excessively subjective belief in the threshold variable decision and to serve as the proxy in deciding which market should be the price leader. The estimation is performed using Markov chain Monte Carlo methods. Furthermore, several meaningful criteria are introduced to assess the forecasting performance in the conditional covariance matrix. The proposed methodology is illustrated using daily S&P500 futures and spot prices. Copyright © 2010 John Wiley & Sons, Ltd.

KEY WORDS dynamic conditional correlation; generalized autoregressive conditional heteroskedasticity; hedge performance; Markov chain Monte Carlo; value at risk

INTRODUCTION

The vector autoregressive (VAR) model, popularized by Sims (1980), has been used widely and extensively by economists to study the dynamic behavior of economic variables. However, most covariance matrices of financial asset returns are serially correlated, and multivariate generalized autoregressive conditional heteroskedastic (GARCH) models that have been introduced to take care of this problem have become increasingly popular in financial econometrics in the past decade. A number of different multivariate GARCH models have been proposed, including the simplified diagonal VECH model of Bollerslev *et al.* (1988), the BEKK model of Engle and Kroner (1995), the constant conditional correlation (CCC) model of Bollerslev (1990), the factor ARCH model of Engle *et al.* (1990) and the dynamic conditional correlation (DCC) model of Engle (2002). In particular, the DCC-GARCH model is simpler and has successfully solved many practical problems. For example, hedges require estimates of the correlation between the returns on the assets. It is well

*Correspondence to: Chih-Chiang Wu, Department of Finance, Yuan Ze University, 135 Yuan-Tung Road, Chungli, Taoyuan, Taiwan 320. E-mail: chihchiang@saturn.yzu.edu.tw

known that financial market volatility changes over time. If the correlations and volatility are changing, then the hedge ratio should be adjusted to account for the most recent information. In addition, the construction of an optimal portfolio with a set of constraints requires a forecast of the covariance matrix of returns.

Furthermore, multivariate threshold models are widely applicable, including co-integrated systems. A growing body of research in the recent time series literature has concentrated on incorporating nonlinear behavior in conventional linear reduced-form specifications such as autoregressive and moving average models. The motivation for moving away from the traditional linear model with constant parameters has typically come from the observation that many economic and financial time series are often characterized by regime-specific behavior and asymmetric responses to shocks. For such series the linearity and parameter constancy restrictions are typically inappropriate and may lead to misleading inferences about their dynamics.

We next consider the situation in which each linear regime follows an autoregressive process. For instance, we have the well-known threshold autoregressive class of models, the statistical properties of which have been investigated in the early work of Tong (1990) and Tsay (1989). Multivariate threshold VARs are piecewise linear models with different autoregressive matrices in each regime, which is determined by a transition variable (one of the endogenous variables), a delay and a threshold (see Tsay, 1998). They were more recently reconsidered and extended in Hansen (2000) and Caner and Hansen (2001), among others. Hansen and Seo (2002) examined a two-regime vector error correction model with a single co-integrated vector and a threshold.

Bayesian methods have also recently become popular to researchers in econometrics. The VAR model usually has a large number of parameters, which are often estimated by means of maximum likelihood or least squares. In the threshold VAR, however, finite-sample frequentist analysis of the nonlinear functions is difficult. For instance, for some distributions of data, the maximum likelihood estimation (MLE) does not have an analytical form or simply does not exist, or in some applications of VAR models nonlinear functions of VAR parameters are the focus of research. The difficulties faced in the frequentist approach to VAR inference can be circumvented by the Bayesian approach, which combines information from observations with researchers' priors. When the objective of the model is to forecast, the Bayesian approach is more satisfactory. This approach consists of imposing prior restrictions on the VAR model parameters. The estimates of the model parameters are obtained by combining the prior belief and the likelihood, so that more accurate forecasts can be achieved (see Bauwens *et al.*, 1999; Vrontos *et al.*, 2003; Osiewalski and Pipień, 2004; So *et al.*, 2005).

In this paper, we present a robust threshold VAR (or VECM)-DCC-GARCH model and use the Metropolis-Hastings (MH) algorithm and the Gibbs sampling algorithm to estimate the parameters simultaneously. Our model extends existing approaches by admitting thresholds in conditional means, conditional volatilities and correlations of multivariate time series. Such an extension allows us to account for rich asymmetric effects and dependencies of conditional means, volatilities and correlations, as they are often encountered in practical financial applications. In addition, we use the concept of Chen and So (2006) to define the threshold variables as the linear combination of endogenous variables. This setting can eliminate excessively subjective belief in threshold variable decision. Besides, the weight coefficient can serve as the proxy in deciding which market is the price leader and which market is the price follower. Finally, the threshold values in our model are not fixed *ex ante* but are estimated from the data, together with all other parameters in the model.

We investigate the empirical performance of our model in the S&P500 futures and spot markets. Our study attempts to use the posterior odds ratio and Bayes factors as a formal tool for making comparisons between competing models. We reduce our testing problem to a Bayesian model

selection problem. We can then select the model with a higher posterior odds ratio. We also present the performance comparison results of the one-step-ahead forecast in the conditional covariance matrix. The forecast results are assessed by several criteria which include the views of statistical loss and risk managers.

Based on the estimation results, we find that the asymmetric dynamic structure is obvious in the dynamic relationship between the S&P500 futures and spot markets. We also detect that the S&P500 futures market is the price leader between the S&P500 futures and spot markets. Furthermore, based on several in-sample and out-of-sample performance measures in the conditional covariance matrix prediction, we find that the threshold model outperforms the linear model across most measurement criteria.

The rest of the paper is organized as follows. In the next section, a robust multivariate threshold vector autoregressive is introduced. Then, the Bayesian approach is specified including the setting of the priors, and then the conditional posterior distributions for relevant parameters are derived. In addition, the Markov chain Monte Carlo (MCMC) simulation method and implementation algorithm are taken into consideration. We then illustrate the empirical applications and the model performance comparisons. Finally brief conclusions are given.

THRESHOLD VAR-DCC-GARCH MODEL

Let Y_t be a K -dimensional time series with threshold variable z_{t-d} , so the threshold VAR-DCC-GARCH model is given by the following equations:

$$Y_t = \sum_{g=1}^G \left[\left(\Phi_0^{(g)} + \sum_{l=1}^{L_g} \Phi_l^{(g)} Y_{t-l} \right) \cdot I(r_{g-1} \leq z_{t-d} < r_g) \right] + \epsilon_t \tag{1}$$

$$\epsilon_t | \mathfrak{S}_{t-1} \sim N(0, H_t) \tag{2}$$

$$H_t = D_t R_t D_t \tag{3}$$

$$D_t^2 = \sum_{g=1}^G \left(\text{diag} \{ \omega_i^{(g)} \} + \sum_{p=1}^{m_g} \text{diag} \{ \alpha_{ip}^{(g)} \} \circ \epsilon_{t-p} \epsilon'_{t-p} + \sum_{q=1}^{n_g} \text{diag} \{ \beta_{iq}^{(g)} \} \circ D_{t-q}^2 \right) \cdot I(r_{g-1} \leq z_{t-d} < r_g) \tag{4}$$

$$R_t = \text{diag} \{ Q_t \}^{-1/2} Q_t \text{diag} \{ Q_t \}^{-1/2} \tag{5}$$

$$Q_t = \sum_{g=1}^G \left((u' - A^{(g)} - B^{(g)}) \circ \bar{Q}^{(g)} + A^{(g)} \circ \eta_{t-1} \eta'_{t-1} + B^{(g)} \circ Q_{t-1} \right) \cdot I(r_{g-1} \leq z_{t-d} < r_g) \tag{6}$$

where $z_t = \sum_{k=1}^K w_k y_{kt}$, $0 \leq w_k \leq 1$, $\sum_{k=1}^K w_k = 1$. $\Phi_0^{(g)}$ is the intercept vector, the $\Phi_l^{(g)}$ for $l = 1, \dots, L_g$ are regression coefficient matrices in the g th regime, $L_g \in \mathbb{N}$ is the lag of VAR in the g th regime, ϵ_t is an innovation term, and $d \in \mathbb{N}$ is the threshold lag of the model with maximum delay d_0 . \mathfrak{S}_{t-1} is the information set up to $t - 1$, and H_t is the time-varying covariance matrix with elements $h_{ii,t}$ and $h_{ij,t}$, $i = 1, \dots, K$, $j = i + 1, \dots, K$. R_t is a time-varying correlation matrix, and D_t is a time-varying diagonal matrix with element

$\sqrt{h_{ii,t}} \cdot \omega_i^{(g)} > 0$, $\alpha_{ip}^{(g)} \geq 0$, $\beta_{iq}^{(g)} \geq 0$, $\sum_{p=1}^{m_g} \alpha_{ip}^{(g)} + \sum_{q=1}^{n_g} \beta_{iq}^{(g)} < 1$, \circ is the Hadamard matrix product operator, $\boldsymbol{\eta}_t = \mathbf{D}_t^{-1} \boldsymbol{\varepsilon}_t$ is the vector of standardized errors, $\bar{\mathbf{Q}}^{(g)}$ is the unconditional correlation matrix of $\boldsymbol{\varepsilon}_t$ in the g th regime, and the parameters $\mathbf{A}^{(g)}$ and $\mathbf{B}^{(g)}$ are symmetric and positive semidefinite matrices. To ensure that \mathbf{Q} is positive semidefinite, we also restrict $(\mathbf{u}' - \mathbf{A}^{(g)} - \mathbf{B}^{(g)})$ being positive semidefinite. In addition, the threshold values r_g must satisfy $-\infty = r_0 < r_1 < \dots < r_G = \infty$, and thus the intervals $[r_{g-1}, r_g), j = 1, \dots, G$, form a partition of the space of z_{t-d} .

The threshold variable z_{t-d} is defined by a weighted average of y_{it-d} and this can be viewed as an extension of Tsay (1998) and Brooks (2001), who set the threshold variable to be a specific endogenous variable, y_{it-d} . We think that this setting of the threshold variable may have the following advantages and economic meanings. First, when the y_{it-d} are the returns of different markets, we can regard w_k and z_{t-d} as the weight and the return of a portfolio without short sales, respectively, and the dynamic structure or leverage effect may be influenced by the portfolio return. That is, the structural change in the markets may rely on the global economic conditions instead of a specific market condition. Second, the setting of the threshold variable can eliminate excessively subjective beliefs in the threshold variable decision and allow the data to choose a more appropriate z_{t-d} by estimating the weights, w_k . Third, the weights w_k can reflect the relative significance of each endogenous variable y_{it-d} . This cannot only govern the time series behavior of Y_t but from it we can also find which market is the price leader and which markets are price followers.

Under the assumption of conditional normality for the error process in equation (2), the likelihood function for the parameters can be expressed as

$$L(\mathbf{y}|\boldsymbol{\theta}) = \prod_{t=P+1}^T (2\pi)^{-\frac{K}{2}} |\mathbf{H}_t|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \boldsymbol{\varepsilon}_t' \mathbf{H}_t^{-1} \boldsymbol{\varepsilon}_t\right\} \tag{7}$$

$$= \prod_{t=P+1}^T (2\pi)^{-\frac{K}{2}} |\mathbf{D}_t \mathbf{R}_t \mathbf{D}_t|^{-\frac{1}{2}} \exp\left\{-\frac{1}{2} \boldsymbol{\varepsilon}_t' (\mathbf{D}_t \mathbf{R}_t \mathbf{D}_t)^{-1} \boldsymbol{\varepsilon}_t\right\}$$

where $\boldsymbol{\theta}$ is the set of all parameters, $P = \max(L_1, \dots, L_G, m_1, \dots, m_g, n_1, \dots, n_g, d_0)$, and $\boldsymbol{\varepsilon}_t$, \mathbf{D}_t , and \mathbf{R}_t obey equations (1), (4), and (5), respectively.

However, when the variables Y_t in the model are integrated and of order one or more, performing the estimation by means of equation (1) is subject to the hazard of regressions involving nonstationary variables. In addition, Engle and Yoo (1987) have also argued that, in the presence of cointegration, a VAR model with an error correction mechanism should outperform a VAR over a longer forecasting horizon. Therefore, by taking into account the explicitly long-run equilibrium relationship, the mean equation (1) is modified to the threshold vector error correction model, which can be written as

$$\Delta Y_t = \sum_{g=1}^G \left[\left(\mathbf{a}^{(g)} \mathbf{b}' Y_{t-1} + \boldsymbol{\Phi}_0^{(g)} + \sum_{l=1}^{P_g} \boldsymbol{\Phi}_l^{(g)} \Delta Y_{t-l} \right) \cdot I(r_{g-1} \leq z_{t-d} < r_g) \right] + \boldsymbol{\varepsilon}_t \tag{8}$$

where Δ denotes the difference operator, $z_t = \sum_{k=1}^K w_k \cdot \Delta y_{kt}$, \mathbf{b} is the $K \times \gamma$ full rank matrix of co-integrating vectors, $\mathbf{a}^{(g)}$ is the $K \times \gamma$ full rank matrix of coefficients associated with the error

correction terms, and the value γ determines the number of co-integrating relationships ($\gamma < K$). To Avoid the identification problem, we impose the so-called linear normalization where $\mathbf{b} = (\mathbf{I}_\gamma \mathbf{b}_o)'$. Thus the likelihood function for the parameters is similar to equation (7) except that $\boldsymbol{\varepsilon}_t$ must follow equation (8).

BAYESIAN INFERENCE AND MCMC IMPLEMENTATION

In this section, we first explain why the Bayesian method is used in this paper to analyze the threshold VAR (or VECM)-DCC-GARCH models. The reason is that by using the maximum likelihood method it is difficult to estimate the parameters in the threshold VAR-DCC-GARCH models. The main problems are the large number of parameters to be estimated and the difficulty of estimation due to the positive definiteness restrictions of the covariance matrix. This will therefore result in unstable estimates. In addition, due to the unknown threshold variable z_t in this paper, we are prevented from implementing a two-step estimation procedure similar to that considered by Tsay (1998) and Brooks (2001). Even if the threshold variable is known, Tsay (1998) showed that the asymptotic properties for the threshold lag, d , and the threshold parameter, r_g , are hard to infer. Thus, to deal with the above unfeasible procedure by using the maximum likelihood method, we extend the Bayesian method using MCMC techniques introduced by Chen and So (2006).

The implementation of the Bayesian analysis depends on a willingness to assign probability distributions not only to the data variable y but also to all unknown parameters. Consider a situation in which absolutely weak previous subjective information is known about the phenomenon of interest, so as to mitigate frequentist criticisms of intentional subjectivity. In this paper, we choose non-informative or weakly informative priors for most parameters to interject the least amount of prior knowledge. The specification of priors is listed below.

First, we assume that the discrete uniform prior for the threshold lag parameter d with maximum delay d_0 can be written as $\pi(d) = 1/d_0$, $d = 1, \dots, d_0$, and does not favor any one of the candidate d values over any other. Since the weighted vector $\mathbf{w} = (w_1, \dots, w_K)$ relies on d , the conditional prior of \mathbf{w} given d , $\pi(\mathbf{w}|d)$ is assumed to be a symmetric Dirichlet distribution, $\mathbf{w}|d \sim D(\delta, \dots, \delta)$, where the hyper-parameter $\delta > 0$. Similarly, z_{t-d} depends on d and \mathbf{w} , and so we assume the conditional prior of every threshold parameter, r_g , for $g = 1, \dots, G - 1$, to be a continuously bounded uniform distribution, $\pi(r_g|d, \mathbf{w}) = 1/(r_{\text{up}}^{(g)} - r_{\text{low}}^{(g)})$, $r_{\text{low}}^{(g)} \leq r_g \leq r_{\text{up}}^{(g)}$. The lower and upper bounds of the threshold parameters are employed to ensure that at least $\tau\%$ of the observations are in each regime. In addition, τ depends on the number of observations. When the sample size is small, a higher τ is recommended. The purpose of this setting is to make the parameter estimates more efficient and more reliable.

We subsequently divide the parameters in each regime into four independent blocks, which are the parameters in the VAR model (1, 8), the error correction term (8), the volatility process (4), and the dynamic correlation procedure (6), and we assume that the priors are independent between any two regimes. For the priors of the VAR parameters, Litterman (1980) and Kinal and Ratner (1986) have indicated that VAR sometimes suffers from overparameterization. The requirement that a large number of coefficients in VAR be estimated often leads to large standard errors for inferences and forecasts. The imposition by the Bayesian VAR of some prior restrictions on parameters will usually provide more accurate forecasts. In this paper, we adopt Litterman's (1980) Minnesota prior for VAR parameters and make some appropriate modifications. For convenience, we define $\boldsymbol{\Phi}^{(g)} = [\boldsymbol{\Phi}_0^{(g)}, \boldsymbol{\Phi}_1^{(g)}, \dots, \boldsymbol{\Phi}_{L_g}^{(g)}]'$, which is a $K \times (1 + K \cdot L_g)$ matrix, and consider vectorized $\boldsymbol{\Phi}^{(g)}$, $\text{vec}(\boldsymbol{\Phi}^{(g)})$, which

we denote by $\phi^{(g)}$. We assume that the vector $\phi^{(g)}$ follows a multivariate normal distribution with zero mean and diagonal covariance matrix Σ_ϕ . That is, we believe in advance that the unconditional mean and short-run dynamics center around zero, and that investors are unable to earn excess returns based on this short-run dynamic relationship. In addition, the priors are made independently across elements of $\phi^{(g)}$, and the standard deviation of the coefficient $\phi_{ijl}^{(g)}$, which is an element of $\phi^{(g)}$ and describes how variable i is affected by variable j of lag l in the g th regime, is given by

$$S(\phi_{ijl}^{(g)}) = \begin{cases} \frac{\lambda}{l} & \text{if } i = j \\ \frac{\eta\lambda}{l} \cdot \frac{\tau_i}{\tau_j} & \text{if } i \neq j \end{cases} \tag{9}$$

where the hyper-parameter λ controls the tightness of beliefs on $\phi^{(g)}$, τ_i/τ_j is a correction for the scale of series i compared with series j , and the restriction $0 < \eta < 1$ implies that the series are more likely to be influenced by their own lags than by the lags of other series.

The above model requires that we choose specific values for the hyper-parameters λ , τ_i , τ_j , and η . The correction term τ_i/τ_j is used in modifying the inconsistency in variation of each series variable. While in principle these should be chosen on the basis of a priori reasoning or knowledge, we will in practice follow Litterman (1986) in choosing these as the sample standard deviations of residuals from univariate autoregressive models that fit the individual series in the sample. For the remaining hyper-parameters, λ is commonly set from 0.1 to 0.9, and η ranges from 0.2 to 0.5 (see Litterman, 1986; Doan, 1990). In addition, Villani (2001) showed that the selection of these two parameters is not sensitive to the forecasting results. We therefore use $\lambda = 0.5$ and $\eta = 0.4$ in our empirical analysis. We also set the standard deviation of the intercept coefficient to 1 to employ a more diffuse prior.

For the prior on the long-term structure (i.e., the error correction term) in each regime, we follow Geweke (1996) to choose uniform priors for both the matrix of co-integrating vectors $b^{(g)}$ and the associated weighting matrix $a^{(g)}$, and the prior can be written as $\pi(a^{(g)}, b^{(g)}) \propto 1$. Furthermore, a uniform prior with some restrictions is assumed for the parameters of the GARCH and is written as

$$\pi(\omega^{(g)}, \alpha^{(g)}, \beta^{(g)}) \propto \prod_{i=1}^K I\left(\omega_i^{(g)} > 0; \alpha_{ip}^{(g)} \geq 0; \beta_{iq}^{(g)} \geq 0; \sum_{p=1}^{m_g} \alpha_{ip}^{(g)} + \sum_{q=1}^{n_g} \beta_{iq}^{(g)} < 1\right) \tag{10}$$

where $I(\cdot)$ is the indicator function, which takes on a value of unity if the constraint holds and zero otherwise, and $\omega^{(g)} = (\omega_1^{(g)}, \dots, \omega_K^{(g)})$, $\alpha^{(g)} = (\alpha_{i1}^{(g)}, \dots, \alpha_{im_g}^{(g)}, \alpha_{K1}^{(g)}, \dots, \alpha_{Km_g}^{(g)})$, and $\beta^{(g)} = (\beta_{i1}^{(g)}, \dots, \beta_{im_g}^{(g)}, \dots, \beta_{K1}^{(g)}, \dots, \beta_{Km_g}^{(g)})$. Finally, we also choose a uniform prior for the dynamic correlation structure

$$\pi(A^{(g)}, B^{(g)}, \bar{Q}^{(g)}) \propto I(\Upsilon) \tag{11}$$

where Υ is the set of $(A^{(g)}, B^{(g)}, \bar{Q}^{(g)})$, which must satisfy the requirement that $A^{(g)}$, $B^{(g)}$, and $(u' - A^{(g)} - B^{(g)})$ are symmetric and positive semidefinite matrices, and $\bar{Q}^{(g)}$ is a form of the correlation coefficient matrix. Therefore, the prior of all unknown parameters in the threshold VAR-DCC-GARCH model can be expressed as

$$\pi(\boldsymbol{\theta}) \propto \pi(d)\pi(\mathbf{w}|d)\left(\prod_{g=1}^{G-1}\pi(r_g|d, \mathbf{w})\right) \cdot \left(\prod_{g=1}^G\pi(\boldsymbol{\phi}^{(g)})\pi(\boldsymbol{\omega}^{(g)}, \boldsymbol{\alpha}^{(g)}, \boldsymbol{\beta}^{(g)})\pi(\mathbf{A}^{(g)}, \mathbf{B}^{(g)}, \bar{\mathbf{Q}}^{(g)})\right) \tag{12}$$

Bayesian inference regarding the parameter vector $\boldsymbol{\theta}$ conditional upon the data matrix \mathbf{y} is constructed through the posterior density $p(\boldsymbol{\theta}|\mathbf{y})$. Using Bayes' theorem, the posterior density is formed by the prior density $\pi(\boldsymbol{\theta})$ and the likelihood $L(\mathbf{y}|\boldsymbol{\theta})$, and it can be expressed as

$$p(\boldsymbol{\theta}|\mathbf{y}) = \frac{L(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})}{\int L(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta})d\boldsymbol{\theta}} \propto L(\mathbf{y}|\boldsymbol{\theta})\pi(\boldsymbol{\theta}) \tag{13}$$

Therefore, the optimal Bayes estimator of $\boldsymbol{\theta}$ under quadratic loss is simply the posterior mean, which is

$$\hat{\boldsymbol{\theta}} = E(\boldsymbol{\theta}|\mathbf{Y}) = \int \boldsymbol{\theta}p(\boldsymbol{\theta}|\mathbf{y})d\boldsymbol{\theta} \tag{14}$$

However, for many realistic problems, the posterior distribution $p(\boldsymbol{\theta}|\mathbf{y})$ may not have an analytically tractable form, particularly for high dimensions, and so calculating the posterior mean is a difficult task. In fact, to settle our major problems, we can use numerical or asymptotic methods to compute the approximate posterior mean for the full Bayesian model. Because the posterior density has a very high dimension and is only known up to a constant, in this paper we adopt the MCMC sample algorithm as our tool for this purpose.

We will therefore use Bayes factors to select the appropriate order of the VAR process and to choose between linear ($G = 1$) and nonlinear ($G \geq 2$) versions of our model. When comparing any two competing parametric Bayesian models (M_i, M_j) for the same data matrix \mathbf{y} , the Bayes factor (BF) can be calculated based on the marginal likelihood concept. In general terms, by letting $\boldsymbol{\theta}_j$ be the appropriate set of parameters under model M_j , the marginal likelihood can be written as

$$p(\mathbf{y}|M_j) = \int_{\boldsymbol{\theta}_j} p(\mathbf{y}|\boldsymbol{\theta}, M_j)p(\boldsymbol{\theta}|M_j)d\boldsymbol{\theta}_j \tag{15}$$

where $p(\mathbf{y}|\boldsymbol{\theta}, M_j)$ and $p(\boldsymbol{\theta}|M_j)$ are the sampling density function and the prior density function, respectively.

For the Bayesian model selection, we can determine the posterior odds ratio (POR) of M_i against M_j by the Bayes factors $B_{ij} = p(\mathbf{y}|M_i)/p(\mathbf{y}|M_j)$ and the prior odds ratio $p(M_i)/p(M_j)$, and it can be expressed as

$$POR_{ij} = \frac{p(M_i|\mathbf{y})}{p(M_j|\mathbf{y})} = \frac{p(M_i)}{p(M_j)}B_{ij} \tag{16}$$

If there is an absence of a prior preference for either model, i.e., $p(M_i) = p(M_j) = 1/2$, the Bayes factors can be interpreted as a measure of the extent to which the data support M_i over M_j . When B_{ij}

> 1 the data prefer M_i over M_j , and when $B_{ij} < 1$ the data favor M_j over M_i . Thus, in this paper, we can compare any two threshold VAR-GARCH time series models of different orders and regimes by computing Bayes factors as the ratio of the marginal likelihood concept evaluated along the route of Chib (1995).

EMPIRICAL APPLICATIONS

Data description

The models are applied to daily data on the S&P500 index and S&P500 index futures for the period 3 January 1995 to 31 October 2007. The sample size for each stock market is 3221. Both spot and futures prices are collected from *TICK DATA*. For S&P500 index futures, we switch to a new contract as the contract's maturity approaches in order to construct a continuous futures contract series. To avoid thin markets and expiration effects, we roll over to the next nearest contract at least one week prior to the expiration of the current contract. The daily spot and futures returns are calculated as the differences in the logarithms of daily price indices multiplied by 100. In addition, intraday 5-minute prices are used to construct the series of realized covariances, as was the case in similar related studies in the past.

The descriptive statistics for S&P500 spot and futures returns are summarized in Table I. The statistics reported are the sample mean, standard deviation, maximum, minimum, Jarque–Bera (JB) statistics, and the Ljung–Box (LB) statistics for the return and the square return series. The standard deviation of futures returns is larger than that of the spot returns, indicating that the futures market is more volatile than the spot market. Both spot and futures returns are negatively skewed and present concerns for excess kurtosis. The JB test statistics provide clear evidence that reject the null

Table I. Descriptive statistics for S&P500 futures and spot returns

	r_{Ft}	r_{St}
Mean (%)	0.038	0.038
SD (%)	1.110	1.069
Skewness	-0.181	-0.138
Kurtosis	6.882	6.480
Max. (%)	5.814	5.308
Min. (%)	-7.706	-7.112
JB	2040.496	1635.338
ADF	-58.831	-57.796
$Q(12)$	28.694	31.957
$Q^2(12)$	902.800	943.49
Johansen test	84.926	

Note: This table reports the descriptive statistics for daily S&P500 futures and spot returns for the sample period from 3 January 1995 to 31 October 2007. r_{Ft} and r_{St} refer to S&P500 futures and spot returns, respectively. Daily returns are calculated by $100 \times (\ln(P_t) - \ln(P_{t-1}))$. The values in rows JB, ADF and Johansen are statistics of the Jarque–Bera normality test, the augmented Dickey–Fuller unit root test, and the Johansen co-integration test, respectively. $Q(12)$ and $Q^2(12)$ report the Ljung–Box (LB) portmanteau test statistics including 12 lags for the return and square return series. The critical values at the 5% level of the JB, ADF, Johansen, and LB statistics are 5.991, -2.862, 15.495, and 21.026, respectively.

hypothesis of normality for each returns series and this is mainly due to the presence of skewness and excess kurtosis. The augmented Dickey–Fuller unit root test results suggest that both return series are stationary. The LB Q statistics indicate possible serial correlations for both returns series. Moreover, the Q^2 statistics suggest that there are autoregressive conditional heteroskedasticity effects on each returns series. The Johansen co-integration test indicates a long-run equilibrium relationship between S&P500 spot and futures indices.

Goodness-of-fit results

In this section, we apply the bivariate threshold VECM-DCC-GARCH model to the S&P500 futures and spot markets. We also use two different time lengths (5 years and 10 years) to perform model comparisons. In addition, we simplify our model by assuming that $L_1 = \dots = L_G = L$, and the GARCH(1,1) model is considered to be a parsimonious model that is found to be appropriate in most applications. The matrix parameters $A^{(g)}$ and $B^{(g)}$ in equation (6) are also reduced to scale parameters. Thus, to ensure that Q is positive semidefinite, we restrict $A^{(g)} + B^{(g)} < 1$, for $g = 1, \dots, G$. Therefore, we only need to choose proper L and G based on the Bayes factor criterion.

To implement our MCMC sampling scheme, we carry out 30,000 iterations, which are performed with the first 10,000 burn-in iterations discarded, to reach the convergence of every parameter. The logarithmic marginal likelihood, the logarithmic Bayes factors, which are on the basis of a simple linear VECM(1)-DCC-GARCH(1, 1) model, and the overall ranking of models for different L and G , are shown in Table II. For the period 2000–2004, Table II shows that the best model to describe the dynamic relationship between the S&P500 futures and spot markets is the two-regime threshold VECM(3)-DCC-GARCH(1, 1) model with a logarithm of the marginal likelihood value of -1926.714 . If we compare the best model with the linear VECM(1)-DCC-GARCH(1, 1) model, i.e., a model in which there is no asymmetric effect on the mean, variance, and correlation equations, then the difference in the logarithmic marginal likelihood is $\ln(B_{6,1}) = 69.606$, which yields a Bayes factor of $B_{6,1} = 1.696 \times 10^{30}$. This means the two-regime threshold VECM(3)-DCC-GARCH(1, 1) model is 1.696×10^{30} times more likely than the linear VECM(1)-DCC-GARCH(1, 1) model. Table II also indicates that the two-regime threshold VECM(3)-DCC-GARCH(1, 1) model is more satisfactory than the other competitors considered here, regardless of whether 5-year or 10-year data are used. Generally speaking, the results in Table II show that models which consider asymmetric effects and longer lags have larger logarithms of marginal likelihood. This suggests that there are asymmetric effects on the mean, covariance, or error correction processes in regard to the dynamic relationship between the S&P500 futures and spot markets. In addition, the logarithmic marginal likelihood of the model with longer lags is larger than that with shorter lags, suggesting that taking longer lags into consideration will benefit the model's explanatory power.

Bayesian estimation results

In this section, we will illustrate the Bayesian estimation results for the period 1995–2004. The linear and threshold VECM(3)-DCC-GARCH(1, 1) models estimation results for the mean equation (8) and variance–covariance matrix equations (4–6) for the dynamic relationship between S&P500 futures and spot markets are presented in Table III. In the three-regime threshold model, the feedback effects between each pair of S&P500 futures and spot markets are observed for all kinds of market conditions (i.e., downside, neutral, and upside markets). That is, lagged spot (futures) returns help to predict current futures (spot) returns. In addition, most lagged spot (futures) returns have negative effects on current spot (futures) returns except for the futures returns in the upside market. The futures (spot) returns tend to decrease (increase) when the spread is large in order to restore the

Table II. Logarithms of marginal likelihood and Bayes factors for S&P500 futures and spot markets

Model (M_i)	Number of parameters	2000–2004 (5 years)			1995–2004 (10 years)		
		$\ln[p(y) M_i]$	$\ln(B_{i,t})$	Rank	$\ln[p(y) M_i]$	$\ln(B_{i,t})$	Rank
M_1 , VECM(1)-DCC-GARCH(1, 1)	18	-1996.319	0	9	-3724.953	0	9
M_2 , VECM(2)-DCC-GARCH(1, 1)	22	-1956.541	39.778	5	-3663.752	61.201	6
M_3 , VECM(3)-DCC-GARCH(1, 1)	26	-1948.295	48.025	4	-3633.945	91.007	4
M_4 , 2R-VECM(1)-DCC-GARCH(1, 1)	38	-1960.836	35.483	6	-3680.614	44.338	7
M_5 , 2R-VECM(2)-DCC-GARCH(1, 1)	46	-1941.648	54.671	2	-3614.597	110.355	2
M_6 , 2R-VECM(3)-DCC-GARCH(1, 1)	54	-1926.714	69.606	1	-3589.465	135.488	1
M_7 , 3R-VECM(1)-DCC-GARCH(1, 1)	56	-1990.747	5.573	8	-3710.036	14.916	8
M_8 , 3R-VECM(2)-DCC-GARCH(1, 1)	68	-1976.574	19.745	7	-3649.549	75.404	5
M_9 , 3R-VECM(3)-DCC-GARCH(1, 1)	80	-1942.786	53.534	3	-3627.081	97.871	3

Note: The table shows the number of parameters, logarithms of marginal likelihood, logarithms of Bayes factors, and the ranks of marginal likelihood for several models and two different time lengths. The data are based on the daily S&P500 futures and spot market prices for the sample period from 3 January 1995 to 31 December 2004.

Table III. Bayesian estimation results for the S&P500 futures and spot markets

	One regime				Two regimes				Three regimes			
	$z_{t-1} \leq r_1$		$z_{t-1} > r_1$		$z_{t-1} \leq r_1$		$z_{t-1} > r_1$		$r_1 < z_{t-1} \leq r_2$		$z_{t-1} > r_2$	
	$\Delta \ln F_t$	$\Delta \ln S_t$	$\Delta \ln F_t$	$\Delta \ln S_t$	$\Delta \ln F_t$	$\Delta \ln S_t$	$\Delta \ln F_t$	$\Delta \ln S_t$	$\Delta \ln F_t$	$\Delta \ln S_t$	$\Delta \ln F_t$	$\Delta \ln S_t$
<i>Mean equations</i>												
ϕ_0	0.031 (0.033)	0.109 (0.030)**	-0.209 (0.097)*	-0.175 (0.106)	0.008 (0.030)	0.088 (0.043)	-0.237 (0.161)	-0.175 (0.147)	-0.025 (0.027)	0.103 (0.043)**	0.529 (0.162)**	0.629 (0.155)**
ϕ_{F1}	-0.222 (0.045)**	0.388 (0.041)**	-0.201 (0.069)**	0.442 (0.062)**	-0.354 (0.049)**	0.234 (0.053)**	-0.194 (0.084)*	0.475 (0.084)**	-0.368 (0.090)**	0.243 (0.083)**	-0.025 (0.077)	0.493 (0.060)**
ϕ_{F2}	-0.128 (0.026)**	0.199 (0.025)**	-0.105 (0.107)	0.219 (0.102)*	-0.159 (0.048)**	0.196 (0.043)**	0.064 (0.058)	0.383 (0.052)**	-0.111 (0.056)*	0.244 (0.054)**	-0.379 (0.071)**	0.012 (0.067)
ϕ_{F3}	-0.158 (0.043)**	0.015 (0.039)	-0.074 (0.129)	0.110 (0.111)	-0.129 (0.044)**	0.048 (0.034)	0.309 (0.112)**	0.479 (0.101)**	-0.136 (0.047)**	0.050 (0.050)	-0.646 (0.147)**	-0.431 (0.146)**
ϕ_{S1}	0.247 (0.050)**	-0.358 (0.047)**	0.049 (0.043)	-0.591 (0.047)**	0.404 (0.048)**	-0.197 (0.053)**	0.031 (0.079)	-0.633 (0.080)**	0.407 (0.070)**	-0.210 (0.067)**	-0.193 (0.058)**	-0.756 (0.056)**
ϕ_{S2}	0.131 (0.033)**	-0.187 (0.033)**	0.044 (0.098)	-0.261 (0.098)**	0.150 (0.049)**	-0.201 (0.042)**	-0.155 (0.068)*	-0.455 (0.062)**	0.123 (0.061)**	-0.228 (0.061)**	0.324 (0.080)**	-0.063 (0.075)
ϕ_{S3}	0.130 (0.043)**	-0.038 (0.038)	0.006 (0.133)	-0.166 (0.116)*	0.134 (0.046)**	-0.046 (0.038)	-0.393 (0.118)**	-0.549 (0.108)**	0.168 (0.048)**	-0.025 (0.050)	0.664 (0.174)**	0.452 (0.175)**
a	-0.821 (0.151)**	0.838 (0.127)**	-0.451 (0.428)	0.376 (0.472)	-0.786 (0.173)**	0.818 (0.155)**	-0.478 (0.351)	0.402 (0.503)	-0.690 (0.264)**	0.816 (0.138)**	-0.293 (0.490)	0.268 (0.435)
b	-1.007 (0.005)**		-1.006 (0.006)**				-1.011 (0.005)**					
<i>Threshold parameters</i>												
d	$1^{\#}$		$1^{\#}$				$1^{\#}$					
w_1	0.671 (0.180)**		0.671 (0.180)**				0.758 (0.076)**					
r_1	-0.601 (0.033)**		-0.601 (0.033)**				-0.620 (0.022)**					
r_2							1.301 (0.019)**					
<i>Variance-covariance equations</i>												
ω	0.033 (0.005)**	0.029 (0.005)**	0.342 (0.038)**	0.312 (0.035)**	0.034 (0.005)**	0.025 (0.004)**	0.332 (0.037)**	0.297 (0.034)**	0.044 (0.006)**	0.030 (0.005)**	0.193 (0.088)*	0.183 (0.079)**
α	0.075 (0.007)**	0.085 (0.009)**	0.069 (0.010)**	0.080 (0.013)**	0.056 (0.011)**	0.064 (0.012)**	0.075 (0.013)**	0.090 (0.015)**	0.020 (0.013)	0.044 (0.015)**	0.044 (0.017)**	0.048 (0.016)**
β	0.899 (0.010)**	0.890 (0.012)**	0.855 (0.025)**	0.836 (0.029)**	0.852 (0.010)**	0.853 (0.011)**	0.851 (0.023)**	0.827 (0.025)**	0.846 (0.011)**	0.847 (0.010)**	0.834 (0.052)**	0.832 (0.056)**
A	0.137 (0.016)**		0.139 (0.039)**		0.232 (0.026)**		0.155 (0.040)**		0.265 (0.034)**		0.165 (0.059)**	
B	0.083 (0.056)		0.122 (0.089)		0.114 (0.060)		0.136 (0.099)		0.066 (0.047)		0.582 (0.178)**	
$\bar{\rho}$	0.975 (0.001)**		0.972 (0.003)**		0.976 (0.001)**		0.972 (0.004)**		0.976 (0.001)**		0.910 (0.202)**	

Note: This table shows the parameter estimates for one, two and three regime threshold VECM(3)-DCC-GARCH(1,1) models, respectively, which are based on the daily S&P500 futures and spot market returns for the sample period from 3 January 1995 to 31 December 2004. Numbers in parentheses are standard deviations. Asterisks indicate statistical significance at the * 5% and ** 1% levels. [#] Posterior mode of the threshold lag parameter.

long-run equilibrium relationship only in the neutral market. The evidence suggests that the S&P500 futures (spot) price tends to converge to the spot (futures) price and the effect is more apparent in the neutral market. Furthermore, the structures of the error correction terms are asymmetric for different market conditions.

The weighted coefficient in the threshold variable, w_1 , is significantly larger than 0.5, especially in the three-regime threshold model. This indicates that the S&P500 futures market is the price leader whereas the spot market is the price follower. Furthermore, we find that the threshold values, r_1 and r_2 , are not symmetric and point out the asymmetric dynamic structures between downside and upside markets. For the coefficients of the variance covariance equations, we find that the volatility is most persistent in the downside market for both the futures and spot markets. However, the persistence of the correlation between futures and spot returns has an opposite outcome. In addition, the unconditional correlation coefficients, $\bar{\rho}$, in the downside market is larger than that in the upside market. The short-term effects of shocks on the correlation are apparent in all kinds of market conditions, while the long-term correlation only influences the current correlation in the upside market.

Covariance matrix forecast comparison

In this section, the data from 3 January 2000 to 31 December 2004, consisting of 1256 trading days, are used for in-sample estimation and forecasting performance evaluation in the one-step-ahead conditional covariance matrix. The 702 observations from 3 January 2005 to 31 October 2007 are used for out-of-sample performance evaluation purposes. The out-of-sample forecasting procedure is carried out as follows. The models are estimated 702 times based on 702 samples of 1256 observations. The first sample, starting 3 January 2000 and ending 31 December 2004, is used to forecast the covariance matrix of 3 January 2005 based on the estimated model for the first sample. The forecast of the covariance matrix is generated for 4 January 2005 based on the estimated model for the second sample, starting 4 January 2000 and ending 3 January 2005. These estimation and forecasting steps can be repeated 702 times for the available sample and we produce the 702 one-step-ahead covariance matrix forecasts.

In addition, we present two categories of criteria to measure the forecasting performance of different competitive models. One category is based on the views of the statistical loss function, which is a non-negative function that generally increases as the distance between the actual value and the forecast value increases, and three different types of criteria are adopted here. The other category of performance measure is based on the views of risk managers and two types of criteria are introduced.

Statistical loss performance

Three types of loss functions are introduced as follows (IS denotes in-sample and OS denotes out-of-sample):

$$\text{IS-MAE} = \frac{1}{T} \sum_{t=1}^T |\text{FCM}_t - \text{RCM}_t|, \text{OS-MAE} = \frac{1}{N} \sum_{t=T+1}^{N_{\text{out}}} |\text{FCM}_t - \text{RCM}_t| \quad (17)$$

$$\text{IS-MSE} = \frac{1}{T} \sum_{t=1}^T (\text{FCM}_t - \text{RCM}_t)^2, \text{OS-MSE} = \frac{1}{N} \sum_{t=T+1}^{N_{\text{out}}} (\text{FCM}_t - \text{RCM}_t)^2 \quad (18)$$

$$\begin{aligned} \text{IS-LINEX}(\zeta) &= \frac{1}{T} \sum_{t=1}^T \{\exp[\zeta(\text{FCM}_t - \text{RCM}_t)] - \zeta(\text{FCM}_t - \text{RCM}_t) - 1\} \\ \text{OS-LINEX}(\zeta) &= \frac{1}{N} \sum_{t=T+1}^{N_{\text{out}}} \{\exp[\zeta(\text{FCM}_t - \text{RCM}_t)] - \zeta(\text{FCM}_t - \text{RCM}_t) - 1\} \end{aligned} \quad (19)$$

where $T = 1256$ is the total number of in-sample observations, $N_{\text{out}} = 702$ is the total number of out-of-sample observations, and ζ is a given parameter to control the asymmetric effect. FCM_t and RCM_t denote the forecast and realized values at time t , respectively.

The first two loss functions are symmetric, which are the mean absolute error (MAE) statistic and the mean square error (MSE) statistic, respectively. The third loss function is asymmetric, which is the linear-exponential (LINEX) loss. When ζ is close to 0, the LINEX loss function is nearly symmetric and is not much different from the MSE statistic. In the LINEX loss function, positive errors are weighed differently from the negative errors when $\zeta \neq 0$. If $\zeta > 0$ ($\zeta < 0$), the LINEX loss function is approximately linear (exponential) for $\text{FCM}_t - \text{RCM}_t < 0$ and exponential (linear) for $\text{FCM}_t - \text{RCM}_t > 0$. This implies that an overestimate (underestimate) needs to be taken more seriously into consideration. More specifically, in all the above cases, a lower loss measure indicates a higher forecasting power.

As a result of the unobservable property of the covariance matrices, here we use intraday 5-minute data to construct the proxies for the daily-realized covariance observations. The concept of the realized volatility has been proposed by French *et al.* (1987) and Andersen *et al.* (2001). The realized volatility is nothing more than the sum of the squared high-frequency returns over a given sampling period. Similarly, we can directly express the realized covariance (RCOV) as

$$\text{RCOV}(t, \Delta) = \sum_{j=1}^{1/\Delta} R(t-1+j \cdot \Delta, \Delta) R(t-1+j \cdot \Delta, \Delta)' \quad (20)$$

where $R(t, \Delta)$ denotes the $K \times 1$ vector of logarithm returns over the $[t - \Delta, t]$ time interval.

A comparison of the results of the forecast performance measures in the conditional covariance matrix between S&P500 futures and spot returns for the different models is presented in Table IV. When looking at the in-sample prediction, we observe that the threshold model yields better performance relative to the linear model for the forecast of covariance between the S&P500 futures and spot returns, $\text{cov}(r_{Ft}, r_{St})$, regardless of which symmetric (MAE and MSE) or asymmetric (LINEX) loss functions are used. The better covariance forecast mainly comes from the improvement in the volatility forecast of the S&P500 futures and spot returns, $\text{var}(r_{Ft})$ and $\text{var}(r_{St})$. The results of the correlation forecasts between S&P500 futures and spot returns, $\text{corr}(r_{Ft}, r_{St})$, for linear or threshold models are similar. For the out-of-sample forecast, two-regime threshold models perform more appropriately for the covariance forecast, while three-regime threshold models have worse forecasting ability than linear models except for the VECM(3)-DCC-GARCH(1, 1) model.

Risk management performance

Predictability in the covariance between two assets' returns, as measured by traditional criteria that focus on the size of the forecast error, does not necessarily imply that an investor can make profits or reduce risk from a trading strategy based on such forecasts. Therefore, we also use the other category of performance measure which is based on the views of risk managers, and two types of

Table IV. In- and out-of-sample covariance matrix forecast comparison of alternative models for the S&P500 futures and spot markets

Model (M_i)	var(r_{F_t})						var(r_{S_t})									
	In-sample			Out-of-sample			In-sample			Out-of-sample						
	MAE	MSE	LIN(1)	LIN(-1)	MAE	MSE	LIN(1)	LIN(-1)	MAE	MSE	LIN(1)	LIN(-1)				
M_1	0.570	1.403	0.388	1909.047	0.237	0.127	0.050	0.158	0.478	0.874	0.258	57.141	0.234	0.118	0.047	0.115
M_2	0.575	1.423	0.375	1770.966	0.224	0.122	0.047	0.156	0.481	0.888	0.252	53.131	0.226	0.113	0.045	0.110
M_3	0.578	1.433	0.373	1774.181	0.214	0.117	0.045	0.147	0.483	0.895	0.252	53.038	0.219	0.111	0.043	0.108
M_4	0.528	1.222	0.359	2013.272	0.218	0.113	0.050	0.094	0.436	0.753	0.237	52.151	0.180	0.090	0.037	0.075
M_5	0.532	1.214	0.370	2198.489	0.223	0.116	0.050	0.099	0.438	0.744	0.243	53.517	0.183	0.093	0.039	0.076
M_6	0.532	1.223	0.390	2130.491	0.211	0.104	0.044	0.085	0.438	0.748	0.248	51.837	0.175	0.084	0.035	0.066
M_7	0.570	1.232	0.331	1819.310	0.293	0.152	0.066	0.123	0.456	0.737	0.220	48.226	0.231	0.112	0.047	0.087
M_8	0.561	1.219	0.338	1931.171	0.268	0.140	0.059	0.119	0.450	0.735	0.226	49.383	0.209	0.105	0.043	0.084
M_9	0.549	1.213	0.348	2152.719	0.203	0.108	0.046	0.095	0.446	0.730	0.230	51.930	0.168	0.087	0.035	0.074

Model (M_i)	corr(r_{F_t}, r_{S_t})						cov(r_{F_t}, r_{S_t})									
	In-sample			Out-of-sample			In-sample			Out-of-sample						
	MAE	MSE	LIN(1)	LIN(-1)	MAE	MSE	LIN(1)	LIN(-1)	MAE	MSE	LIN(1)	LIN(-1)				
M_1	0.045	0.004	0.002	0.002	0.054	0.004	0.002	0.002	0.451	0.736	0.237	16.692	0.208	0.104	0.040	0.117
M_2	0.044	0.004	0.002	0.002	0.053	0.004	0.002	0.002	0.454	0.748	0.230	18.597	0.200	0.101	0.038	0.113
M_3	0.045	0.004	0.002	0.002	0.055	0.004	0.002	0.002	0.456	0.753	0.230	18.698	0.191	0.098	0.036	0.109
M_4	0.045	0.004	0.002	0.002	0.053	0.004	0.002	0.002	0.417	0.625	0.219	10.042	0.176	0.085	0.034	0.075
M_5	0.045	0.004	0.002	0.002	0.054	0.004	0.002	0.002	0.419	0.619	0.224	10.341	0.180	0.088	0.035	0.077
M_6	0.045	0.004	0.002	0.002	0.055	0.004	0.002	0.002	0.420	0.627	0.231	9.676	0.171	0.079	0.032	0.067
M_7	0.045	0.004	0.002	0.002	0.054	0.004	0.002	0.002	0.441	0.623	0.204	9.670	0.235	0.110	0.045	0.090
M_8	0.045	0.004	0.002	0.002	0.055	0.004	0.002	0.002	0.434	0.619	0.208	9.558	0.215	0.103	0.041	0.088
M_9	0.045	0.004	0.002	0.002	0.055	0.004	0.002	0.002	0.428	0.612	0.212	9.212	0.166	0.083	0.032	0.075

criteria are used. One involves calculating the value at risk (VaR) as an evaluation of the estimator. For a two-asset portfolio with δ_1 invested in the first asset and δ_2 in the second asset, the one-step-ahead VaR at time t and at $\alpha\%$, assuming normality, is

$$\text{VaR}_t(\alpha) = -\left[(\delta_1 \hat{r}_{1,t} + \delta_2 \hat{r}_{2,t}) - z_\alpha \cdot \sqrt{\delta_1^2 \hat{h}_{11,t} + \delta_2^2 \hat{h}_{22,t} + 2\delta_1 \delta_2 \hat{h}_{12,t}} \right] \quad (21)$$

where z_α is the right quantile at $\alpha\%$. To compare and evaluate model performances for several different model specifications, we choose the following criteria. By definition, the failure rate (FR) is the proportion of returns (in absolute value terms) that exceed the forecast VaR, i.e. $\text{FR} = (1/T) \sum_{t=1}^T I(\delta_1 r_{1,t} + \delta_2 r_{2,t} < -\text{VaR}_t(\alpha))$, where $I(\cdot)$ is the indicator function. Hence, if the VaR model is correctly specified, the failure rate should be equal to the prespecified VaR level. We also use the Kupiec (1995) LR test to examine whether the model is correctly specified. To test $H_0: f = \alpha$ against $H_1: f \neq \alpha$, the LR statistic is $\text{LR} = -2 \ln(\alpha^N (1 - \alpha)^{T-N}) + 2 \ln((N/T)^N (1 - (N/T))^{T-N})$, where N is the number of VaR violations, T is the total number of observations and f is the theoretical failure rate. Under the null hypothesis, the LR test statistic is asymptotically distributed as $\chi^2(1)$.

The fractions of VaR violations and p -values of the Kupiec (1995) failure rate test for a hedged portfolio with weights $(\delta_{\text{futures}}, \delta_{\text{spot}}) = (-1, 1)$ are reported in Table V. For in-sample data, the p -values for the null hypothesis of the hedged portfolio are all smaller than 0.05 when the linear model is considered. The threshold model performs very well as there are no p -values smaller than 0.05. Thus the switch from the linear model to the threshold model yields a significant improvement in the VaR performance in the hedged portfolio. For the out-of-sample VaR performance comparison, the p -values of the linear and threshold models with the lag of VAR $L = 1$ are all smaller than 0.05. However, for $L = 2$ and $L = 3$, we find that the fractions of the VaR violation based on the threshold model are closer to the prespecified VaR level than those based on the linear model. In view of the in-sample and out-of-sample empirical results, we find that the linear model may sometimes overestimate the VaR, and so using the threshold model to calculate the portfolio's VaR may be more appropriate.

Table V. In- and out-of-sample 5% VaR failure rate results for the S&P500 futures–spot hedged portfolio

Model (M_i)	(A) Fraction of VaR violation		(B) p -value of Kupiec LR test	
	In-sample	Out-of-sample	In-sample	Out-of-sample
M_1	2.070	3.419	0.000	0.042
M_2	1.592	3.846	0.000	0.144
M_3	2.787	3.989	0.000	0.203
M_4	4.459	3.419	0.370	0.042
M_5	4.618	4.131	0.529	0.277
M_6	5.971	4.131	0.125	0.277
M_7	5.573	3.419	0.360	0.042
M_8	5.096	4.131	0.877	0.277
M_9	5.175	4.558	0.777	0.586

Note: This table shows the 5% VaR forecast results of S&P500 futures–spot hedge portfolios for alternative models. Panel (A) is the fraction of VaR violations, and the results of the Kupiec LM test (1995) are shown in panel (B). The in-sample data period extends from 3 January 2000 to 31 December 2004 and the out-of-sample data period extends from 3 January 2005 to 31 October 2007. M_1, M_2, \dots, M_9 refer to the same models as in Table II.

While futures contracts are popular among investors as a class of speculative assets, they are important in the financial markets due to their use as a hedging instrument. Furthermore, hedging with futures contracts may be the simplest method to manage market risk resulting from adverse movements in the prices of various assets. In this section, we assume the hedger attempts to minimize the conditional variance of the spot–futures portfolio. It is well known that the optimal hedge ratio (OHR) is the ratio of the conditional covariance between spot and futures returns over the conditional variance of the futures return. Thus the one-step-ahead forecasts of optimal hedge ratios can then be calculated as

$$HR_t^* = \text{cov}_{t+1|r}(r_s, r_f) / \text{var}_{t+1|r}(r_f) \quad (22)$$

where r_s and r_f are spot and futures returns, respectively. The variance of the estimated optimal hedged portfolio can be characterized as

$$\text{var}(r_{s,t} - HR_t^* \cdot r_{f,t}) \quad (23)$$

To evaluate hedging performance, the typical criterion is based on the percentage variance reduction (PVR) of the hedged portfolio relative to the unhedged position. It can be calculated as

$$\text{PVR}(\%) = \left[1 - \left(\frac{\text{var}(\text{hedged portfolio})}{\text{var}(\text{unhedged portfolio})} \right) \right] \times 100\% \quad (24)$$

When the futures contract completely eliminates risk, PVR = 100 is obtained, otherwise PVR = 0 is obtained when hedging with the futures contract does not reduce risk. Hence a larger PVR indicates better hedging performance.

The in- and out-of-sample hedged portfolio variances and hedging effectiveness of alternative models for the S&P500 futures contract are presented in Table VI. The variances of hedged portfolio returns are calculated under the following 11 alternative models: three linear VECM-DCC-GARCH models with different lag parameters $L(M_1, M_2, M_3)$, six threshold VECM-DCC-GARCH models with different lag parameters L and the number of regimes $G(M_4, \dots, M_9)$, hedging with a constant OHRs estimate using regression methods of returns and the naïve hedge with a hedge ratio of 1 at all times. The results show that the three-regime threshold VECM(3)-DCC-GARCH model has the lowest in-sample hedged portfolio variance, with a 94.587% in-sample variance reduction compared to the variance of the unhedged position. In addition, the in-sample hedging performance of the linear model with $L = 1$ is even worse than that of OLS or a naïve strategy.

However, active hedgers are likely to be more concerned about future hedging performance. Therefore, the comparison of out-of-sample performance is a better way to evaluate our hedging strategy. We find that the ranking of out-of-sample hedging effectiveness is very similar to that of in-sample hedging effectiveness. The two-regime threshold VECM(2)-DCC-GARCH model has a 98.145% out-of-sample variance reduction and outperforms all of the linear dynamic and static hedging models that we considered. Overall, the dynamic hedge with a threshold model has better hedging performance than that with the linear VECM-DCC-GARCH model. This fact may indicate that the threshold model has a superior ability to forecast the optimal hedge ratios.

Table VI. In-sample and out-of-sample hedging effectiveness of alternative models for S&P500 spot and futures markets

Model (M_i)	In-sample			Out-of-sample		
	Variance	Variance reduction	Rank	Variance	Variance reduction	Rank
M_1	0.09259	94.26382	11	0.03069	98.09868	11
M_2	0.09037	94.40135	8	0.03061	98.10363	10
M_3	0.08969	94.44348	7	0.03046	98.11293	8
M_4	0.08908	94.48127	6	0.03015	98.13213	4
M_5	0.08806	94.54446	2	0.02994	98.14514	1
M_6	0.08860	94.51101	5	0.02999	98.14204	2
M_7	0.08808	94.54322	3	0.03010	98.13523	3
M_8	0.08824	94.53331	4	0.03039	98.11726	6
M_9	0.08737	94.58721	1	0.03043	98.11479	7
OLS	0.09045	94.39640	9	0.03023	98.12718	5
Naïve	0.09182	94.31152	10	0.03058	98.10549	9

Note: The table reports the variance of the hedged portfolio, percentage variance reductions, and the ranks of hedging effectiveness for several different models. The data used are daily index futures and spot prices. The in-sample data period extends from 3 January 2000 to 31 December 2004 and the out-of-sample data period from 3 January 2005 to 31 October 2007. M_1, M_2, \dots, M_9 refer to the same models as in Table II.

CONCLUSIONS

We have proposed a robust multivariate VAR-DCC-GARCH model that extends existing approaches by admitting multivariate thresholds in conditional means, conditional volatilities and conditional correlations. In addition, such threshold variables are defined by a weighted average of endogenous variables and the weights are estimated from the data. This threshold setting cannot only enhance the robustness of the model but also has some economic meaning or value. Moreover, the MCMC method is implemented for the Bayesian inference. We have studied the performance of our model in an application to daily S&P500 futures and spot prices.

We develop a Bayesian testing scheme for model selection among several competing models and select the model with a higher posterior probability. We also adopt several criteria, which are based on the views of statistical loss and risk managers, to evaluate the prediction performance of the conditional covariance matrix.

In our real data application we find that estimated conditional volatilities are strongly characterized by both GARCH and multivariate threshold effects. Dynamic correlations are still apparent between the S&P500 futures and spot markets. In addition, the estimation results suggest that the S&P500 futures market is the price leader between the S&P500 futures and spot markets. Based on a comparison of covariance matrix forecasting performance, it is found that the threshold model has better in-sample and out-of-sample forecasting performance relative to the linear model across most measurement criteria.

REFERENCES

- Andersen T, Bollerslev T, Diebold F, Ebens H. 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* **61**: 43–76.
- Bauwens L, Lubrano M, Richard JF. 1999. *Bayesian Inference in Dynamic Econometric Models*. Oxford University Press: New York.

- Bollerslev T. 1990. Modeling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH model. *Review of Economics and Statistics* **72**: 498–505.
- Bollerslev T, Engle R, Wooldridge JM. 1988. A capital asset pricing model with time varying covariances. *Journal of Political Economy* **96**: 116–131.
- Brooks C. 2001. A double-threshold GARCH model for the French Franc/Deutschmark exchange rate. *Journal of Forecasting* **20**: 135–143.
- Caner M, Hansen BE. 2001. Threshold autoregression with a unit root. *Econometrica* **69**: 1555–1596.
- Chen CWS, So MKP. 2006. On a threshold heteroscedastic model. *International Journal of Forecasting* **22**: 73–89.
- Chib S. 1995. Marginal likelihood from the Gibbs output. *Journal of the American Statistical Association* **90**: 1313–1321.
- Chib S, Greenberg E. 1995. Understanding the Metropolis–Hastings algorithm. *American Statistician* **49**: 327–335.
- Doan TA. 1990. *Regression Analysis of Time Series*. VAR Econometrics: Evanston, IL.
- Engle RF. 2002. Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business and Economic Statistics* **20**: 339–350.
- Engle RF, Kroner K. 1995. Multivariate simultaneous GARCH. *Econometric Theory* **11**: 122–150.
- Engle RF, Yoo B. 1987. Forecasting and testing in cointegrated systems. *Journal of Econometrics* **35**: 143–159.
- Engle RF, Ng V, Rothschild M. 1990. Asset pricing with a factor-ARCH covariance structure: empirical estimates for Treasury bills. *Journal of Econometrics* **45**: 213–238.
- French KR, Schwert GW, Stambaugh RF. 1987. Expected stock returns and volatility. *Journal of Financial Economics* **19**: 3–29.
- Geweke J. 1996. Bayesian reduced rank regression in econometrics. *Journal of Econometrics* **75**: 121–146.
- Hansen BE. 2000. Sample splitting and threshold estimation. *Econometrica* **68**: 575–603.
- Hansen BE, Seo B. 2002. Testing for two-regime threshold cointegration in vector error-correction models. *Journal of Econometrics* **110**: 293–318.
- Kinal T, Ratner J. 1986. A VAR forecasting model of a regional economy: its construction and comparative accuracy. *International Regional Science Review* **10**: 113–126.
- Kupiec P. 1995. Techniques for verifying the accuracy of risk management models. *Journal of Derivatives* **3**: 73–84.
- Litterman RB. 1980. *Techniques for forecasting with vector autoregressions*. Working paper, Massachusetts Institute of Technology.
- Litterman RB. 1986. Forecasting with Bayesian vector autoregressions: five years of experience. *Journal of Business and Economic Statistics* **4**: 25–38.
- Osiewalski J, Pipień M. 2004. Bayesian comparison of bivariate ARCH-type models for the main exchange rates in Poland. *Journal of Econometrics* **123**: 371–391.
- So MKP, Chen CWS, Chen MT. 2005. A Bayesian threshold nonlinearity test for financial time series. *Journal of Forecasting* **24**: 61–75.
- Sims C. 1980. Macroeconomics and reality. *Econometrica* **48**: 1–48.
- Tong H. 1990. *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press: New York.
- Tsay RS. 1989. Testing and modeling threshold autoregressive process. *Journal of the American Statistical Association* **84**: 231–240.
- Tsay RS. 1998. Testing and modeling multivariate threshold models. *Journal of the American Statistical Association* **93**: 1188–1998.
- Villani M. 2001. Bayesian prediction with cointegrated vector autoregressions. *International Journal of Forecasting* **17**: 585–605.
- Vrontos D, Dellaportas P, Politis DN. 2003. Inference for some multivariate ARCH and GARCH models. *Journal of Forecasting* **22**: 427–446.

Authors' biographies:

Chih-Chiang Wu is Assistant Professor of Finance at Yuan Ze University. His research interests include forecasting, quantitative finance, and risk management.

Jack C. Lee, an ASA Fellow, is University Chair Professor of Finance at National Chiao Tung University. His research interests include Bayesian inference, financial statistics, forecasting, growth curve, and time series analysis.

Authors' addresses:

Chih-Chiang Wu, Department of Finance, Yuan Ze University, 135 Yuan-Tung Road, Chungli, Taoyuan, Taiwan 32003.

Jack C. Lee, Institute of Finance, National Chiao Tung University, 1001 Ta-Hsueh Road, Hsinchu, Taiwan 30050.