Volatility Analysis via Coupled Wishart Process

Zhong She and Can Wang

Abstract—The study of volatility has been a great concern to people both from academia and industry. Recently, several stochastic approaches such as the Wishart process, have been proposed to model the volatility with strong capturing power and flexibility. However, these kinds of models and their deviations only tackle several variables from only one system. But in real world, individuals and systems are more or less related to each other via explicit or implicit relationships. In this paper, we propose a coupled Wishart process model to capture such couplings when modeling the volatility, in which a linear effect. For this purpose, multivariate volatility models [6] have been proposed based on the univariate ones. Multivariate volatility modeling methods is in great demand. Modeling and predicting volatility is also of vital importance in various applications [2], [3], [4], such as asset pricing, portfolio selection, hedging and risk management. In addition, the burst of several bubbles aggravates the uncertainty of world market and economy. Especially after the economic crisis in 2008, modeling and analyzing volatility is becoming a more and more critical task.

There has been a great number of research [5] on the univariate volatility models, which consider the volatility of just one variable. However, studies on the volatility of a certain financial instrument should also include the influences from other assets to capture the dependency and spill-over effect. For this purpose, multivariate volatility models [6] have been proposed based on the univariate ones. Multivariate volatility models focus on the covariance matrix of a portfolio, and regard that the volatilities evolve together over time, according to specific model setups.

To model the dynamic covariance matrix evolution in multivariate volatility analysis, Wishart process has been adopted in several literatures [3], [7]. As a probabilistic model, Wishart process demonstrates its flexibility and simplicity in modeling the covariance matrix [6]. In a Wishart process, the return of one asset is regarded as an observation of the latent covariance matrix, and conditional independent with others. The latent covariance matrix evolves in a one-order Markov way: only related to the state of its last time.

Nevertheless, under the circumstance of globalization, only considering the influences from other assets is not enough. Outside influences often play an important role than one system itself. Like the world-wide economic crises in history, all of them start from one market and then spread to the rest of whole world. Imagine that a professional investor is managing a portfolio in Hong Kong stock market. He needs a good estimation on the covariance matrix of this portfolio. As the Hong Kong stock market is highly correlated with the U.S. market, taking into account the U.S. market will definitely help him make better estimation of the covariance matrix. However, no work has explicitly and systematically address the coupling relationship across systems or markets for volatility due to its complexity with great challenges. The incomplete or local analysis of volatility will inevitably lead to tentative and less effective learning performance.

Here, we introduce the thought of coupling and include the outside influences from other markets. Both in nature and society, there exist some kinds of coupling relationships between individuals and systems. The research on such couplings and interactions is of great significance in diverse fields, like the coupled behavior analysis [8], the coupled anomaly detection [9], and the coupled attribute analysis [10]. In the example of the Hong Kong investor, a coupled volatility analysis framework can be constructed as in Fig. 1. Within this framework, at each time point, the covariance matrix of a portfolio from Hong Kong market is evolving with that of a similar portfolio from U.S. market.

In this paper, we propose the coupled volatility analysis with linear coupling relationship between the covariance matrices. This relationship is dynamic with time, providing much more flexibility and better accuracy for models. In
summary, the key contributions are listed as follows:

- We propose a coupled Wishart process by considering the interactions between different systems, in which we construct linear coupling relationship with properly designed weights.
- An estimation method based on MCMC, i.e. Gibbs sampling and Metropolis-Hasting sampling, is adopted to learn all the parameters and hidden covariance matrices.
- Rigorous experiments have been conducted to show the superiority of our proposed coupled model.

The remainder of this paper is organized as follows. In Section II, we briefly review the related work. Basic knowledge on Wishart distribution and Wishart process is presented in Section III. Section IV proposes the coupled volatility analysis model, a linear coupling approach is introduced in Section V. The estimation method for this model is then specified in Section VI. Experiment results are shown in Section VII, which prove the effectiveness of our proposed model in capturing the coupling relationship. Finally, the conclusion and future work are presented in Section VIII.

II. RELATED WORK

Volatility analysis is an important research issue that has been broadly studied in econometrics and other communities [11]. The univariate volatility models deal with the volatility of just one variable [5]. However, such methods lack the consideration of the influences from other variables. Thus, univariate volatility analysis has been extended to the multivariate cases by several researchers [12]. Two main approaches for volatility analysis are GARCH models and stochastic models. The GARCH models assume that the volatility is a deterministic function of the past [1], [11], while the stochastic models suppose the volatility follows a random function defined as

$$ y_t | \Sigma_t \sim N(0, \Sigma_t), $$

$$ \Sigma_t^{-1} | \nu, \Sigma_{t-1}^{-1} \sim W_k(\nu, S_{t-1}), $$

where $\Sigma_t$ is the matrix random variable, $\nu$ is a scalar parameter larger than $k$, $W$ is a $k \times k$ symmetric positive definite matrix parameter, and $\Gamma_k(\cdot)$ is the multivariate gamma function defined as

$$ \Gamma_k(\nu/2) = \pi^{k(k-1)/4} \prod_{j=1}^{k} [\nu + 1 - j]/2. $$

B. Wishart Process

The Wishart process [3], [7], [17] is proposed to model the dynamic time series of covariance matrices. The approach introduced by Philipov and Glickman [3] is adopted in this paper. The graphical model for a Wishart process is presented in Fig. 2. The dynamic covariance structure is modeled on Wishart distribution:

$$ y_t \Sigma_t \sim N_k(0, \Sigma_t), $$

$$ \Sigma_t^{-1} | \nu, \Sigma_{t-1}^{-1} \sim W_k(\nu, S_{t-1}), $$

where $S_{t-1}$ is defined as follows,

$$ S_{t-1} = \frac{1}{\nu} (A^{1/2})(\Sigma_{t-1}^{-1})^{\nu/2}(A^{1/2})'. $$

The variables and parameters in the above equations are explained here. $y_t$ is the return vector of $k$ different assets at time $t$, and is set to obey a zero-mean multivariate Gaussian distribution (denoted as $N_k$). So it can be regarded as the return that “surprises” the expected. $\Sigma_t$ is the latent covariance matrix of returns at time $t$, and treated as the volatility for a
portfolio. $\Sigma_t$ and $S_t$ are symmetric positive definite matrices, defined on $\mathcal{M}_k^+ \subset \mathbb{R}^{k \times k}$. $\nu$ is the degree of freedom and is set to be invariant during the whole process. $A \in \mathcal{M}_k^+$ is also a symmetric positive definite matrix parameter, and can be decomposed through a Cholesky decomposition, denoted as $A = (A^{1/2})(A^{1/2})'$. This parameter matrix reveals how each entry of covariance matrix $\Sigma_t$ at time $t$ depends on the entries of covariance matrix $\Sigma_{t-1}$ at time $t-1$. So $A$ can be interpreted as a measure of contemporaneous sensitivity [3]. $d$ is a scalar parameter to measure the overall strength of relationship between previous period and current period. As discussed in [3], we set $d \in (0, 1)$. Besides, $t = 1, 2, \ldots, T$ is the time indicator.

IV. Coupled Volatility Analysis

Considering the interaction of world financial markets, the volatility of one market is greatly influenced by that of other markets. Therefore, constructing a coupled volatility model to analyze the volatility between several different markets is quite essential. Inspired by the research [8] based on the Hidden Markov Model (HMM) and some works on the coupled time series [18], a coupled model on volatility analysis is proposed.

Below, we present a coupled Wishart process to capture the interactive volatilities upon multiple Wishart processes. Within a coupled framework, it is assumed that the covariance matrix is conditioned with not only the previous state of its own Wishart process, but also its “neighbor” covariance matrix at last step from other Wishart processes.

![Fig. 3. An example of the coupled Wishart process.](image)

Take a two-chain coupled Wishart process for example, the graphic model is exhibited in Fig. 3. The two Wishart processes (i.e. $W$ and $W'$) are coupled during the procedure of evolving. As can be observed from Fig. 3, at each time $t$, the covariance matrix $\Sigma_t$ is not only determined by its previous one $\Sigma_{t-1}$, but also depends on $\Sigma'_{t-1}$. Formally, we have the formalization as below.

$$y_t|\Sigma_t \sim N_k(0, \Sigma_t)$$

$$\Sigma_t^{-1}|\nu, \Sigma_{t-1}, (\Sigma'_{t-1})^{-1} \sim W_k(\nu, g(\Sigma_{t-1}, \Sigma'_{t-1}))$$

$$y'_t|\Sigma'_t \sim N_k(0, \Sigma'_t)$$

$$\Sigma'_t^{-1}|\nu', \Sigma_{t-1}, (\Sigma'_{t-1})^{-1} \sim W_k(\nu', g'(\Sigma_{t-1}, \Sigma'_{t-1}))$$

As well presented in Equation (7) and Equation (9), at a new time $t$, the coupled evolvement is passed from last stage $t-1$ by a coupling of covariance matrices, i.e. $g(\Sigma_{t-1}, \Sigma'_{t-1})$ and $g'(\Sigma_{t-1}, \Sigma'_{t-1})$. Here, $g(\cdot)$ and $g(\cdot)'$ are the coupling functions of covariance matrices. How to specifically construct $g(\cdot)$ and $g(\cdot)'$ is another issue, which will be addressed in the next section.

V. A Linear Coupling Approach

In this section, we present a linear coupling model for the coupled volatility analysis and its weights setting, together with their theoretical supports.

A. Linear Model

Based on several coupled frameworks on HMM [15] and linear dynamic systems [18], which consider both inter-coupling and intra-coupling in a linear way, a similar coupled model is constructed accordingly. Below is a coupled system with $Q$ Wishart processes, for each process $W_q$ ($q = 1, 2, \ldots, Q$) at every time point $t$, the evolution of the corresponding covariance matrix is coupled with the other $Q-1$ covariance matrices.

$$W^{(1)} : \Sigma^{(1)}_t \sim W_k(\nu^{(1)}, g^{(1)}_{t-1})$$

$$W^{(2)} : \Sigma^{(2)}_t \sim W_k(\nu^{(2)}, g^{(2)}_{t-1})$$

$$\ldots$$

$$W^{(Q)} : \Sigma^{(Q)}_t \sim W_k(\nu^{(Q)}, g^{(Q)}_{t-1})$$

$$W^{(Q)} : \Sigma^{Q}_{t} \sim W_k(\nu^{Q}, g^{Q}_{t-1})$$

In detail, for the $q$th Wishart process $W^{(q)}$ at time $t$, the evolution of covariance is determined by two parameters: $\nu^{(q)}$ and $g^{(q)}_{t-1}, \nu^{(q)}$ is the parameter that indicates its own property, and set to be invariant for each process. $g^{(q)}_{t-1}$ is a function of the linear combination of all the covariance matrices at time $t-1$, note that it is a linear specification of $g(\cdot)$ in Equation (7) and Equation (9).

The linear combination function $g^{(q)}_{t-1}$ is defined as follows. Equation (11) is based on the property of Wishart process, and its structure is similar to Equation (5). Equation (12) illustrates how exactly the coupling goes. The weights $\{\omega^{(j)}_{t-1}\}$ represents the extent of influences at time $t$ from the $j$th Wishart process. $\alpha$ is a weight that indicates the self-influence and set to be invariant at our current stage.

$$g^{(q)}_{t-1} = \frac{1}{\nu^{(q)}}(A^{1/2}_q)\Sigma^{(q)}_{t-1} - d^{(q)} (A^{1/2}_q)'$$

$$\Sigma^{(q)}_{t-1} = \alpha \cdot \Sigma^{(q)}_{t-1} + (1 - \alpha) \cdot \sum_{j=1}^{Q} \omega^{(j)}_{t-1} \cdot \Sigma^{(j)}_{t-1}$$

The linear structure of coupling relationship on covariance matrices has several advantages. First, linear relationship leads to a clear and easy-to-inference structure, the coupling influences from other systems are modeled with weights. Second, under the linear structure, the self-evolution can...
be easily modeled, which is an very important part in the evolvement \[18\].

Here we take the example from Section I to demonstrate how linear coupling of matrices makes sense. Assume \(\Sigma_t\) and \(\Sigma'_t\) are the covariance matrices of two identical assets portfolio (i.e. manufacturing and consuming) from U.S. and Hong Kong stock markets respectively, we make a linear combination \(C\) of them. The linear combination of these two matrices with weights, corresponds to the linear combination of every entry in the matrices. Any element of the combined matrix, like \(C(1,1)\) indicates the corresponding linear combination of element \(\Sigma_t(1,1)\) from covariance matrix \(\Sigma_t\) and \(\Sigma'_t(1,1)\) from covariance matrix \(\Sigma'_t\).

Here naturally arises another problem: how to acquire a set of reasonable weights? As the coupling objects are not ordinary variables, but positive definite matrices, some related issues are addressed in the next part.

### B. Weights Setting

How to define proper weights for those covariance matrices is an important issue. We define the weights as follows

\[
\omega^{(j)}_t = \frac{||y_t^{(j)} - y_{t-1}^{(j)}||_2}{\sum_{j=1}^{Q} ||y_t^{(j)} - y_{t-1}^{(j)}||_2}
\]

This selection of weights is based on the thought that the violent system tends to exert a greater influence on the quieter ones \[6\]. So the extent of changes in observations at time \(t-1\) determines the weight of evolvement to next time stage.

### C. Theoretical Support

First, we easily prove that a linear combination of covariance matrices is still a symmetric positive definite matrix, which means that the linear structure is a reasonable candidate to capture the coupling relationship.

**Theorem 5.1:** Suppose \(A\) and \(B\) are two symmetric positive definite matrices, \(\omega_1, \omega_2 \in (0,1)\) are two weights. A linear combination of \(A\) and \(B\) is still a symmetric positive definite matrix.

In fact, for any non-zero vector \(x\), we have

\[
x(\omega_1 A + \omega_2 B)x' = \omega_1 \cdot xAx' + \omega_2 \cdot xBx' > 0
\]

According to the definition of symmetric positive definite matrix, the linear combination of them is still a symmetric positive definite matrix.

Next, we demonstrate that a linear combination of covariance matrices \(A\) and \(B\), denoted as \(C\), is still between \(A\) and \(B\) on the space of covariance matrices \(M_{+}^k\). The reason is that we accordingly have \(d(A, C) \leq d(A, B)\) and \(d(B, C) \leq d(A, B)\), detailed proof is available by request. As our target objects are not a vector or a scalar, we apply a distance metric on covariance matrix \[19\].

\[
d(A, B) = \sqrt{\sum_{i=1}^{m} \ln^2 \lambda_i(A, B),}
\]

where \(\lambda_i(A, B)\) is the eigenvalue from \(|\lambda A - B| = 0\). Positivity, symmetry, and triangle inequality of this metric can be found in \[19\].

### VI. Estimation Methods

For such a probabilistic framework with a clear graphic model, Markov chain Monte Carlo (MCMC) \[20\] methods are promising ways to simulate the time series, as the parameters and latent variables can be estimated together. In this section, we propose an MCMC technique to estimate the coupled Wishart process.

As a Bayesian approach, the idea behind MCMC methods is to produce values for specific variables from a known distribution of interest (usually multivariate distribution) by sampling a Markov chain, whose invariant transition distribution is just the target distribution. Unlike the usual problem of acquiring the maximum likelihood of parameter \(\theta\), in this model, the parameter space is augmented to include all the latent variables. After the Markov chain converges to the target distribution, these draws are treated as the samples from marginal posterior densities. Therefore, several statistics such as mean and moment can be approximately calculated.

From the construction of our model, we find that our target posterior distribution has a high-dimension density (see Equation \[16\]), with both scalar and matrix variables. For such a high-dimension posterior, Gibbs sampling \[20\] is a good choice, which produces samples of every element from its conditional probability with all others, then repeats this process while updating the samples according to the conditional probability density function. The most advantage of adopting Gibbs sampling is that it produces high-dimension samples via its own procedures. Besides, a Gibbs sampler accepts all the candidate draws \[20\], which avoids the trouble of choosing a proper proposal distribution in practice.

To implement the Gibbs sampling, we deduce the joint posterior of parameters and latent variables first. The joint posterior function in a coupled Wishart process is proportional to the priors multiplied by the likelihood. Take a two-chain coupled Wishart process as an example, below is the posterior distribution:

\[
p(\nu^{(1)}, \nu^{(2)}, A^{(1)}, A^{(2)}, d^{(1)}, d^{(2)}, \Sigma_1^{1:T}, \Sigma_2^{1:T}, y_1^{1:T}, y_2^{1:T}) \
\propto p(\nu^{(1)}) \cdot p(\nu^{(2)}) \cdot p(A^{(1)}) \cdot p(A^{(2)}) \cdot p(d^{(1)}) \cdot p(d^{(2)}) \cdot \frac{T}{\prod_{i=1}^{T} p(\Sigma_1^{i:T}) \prod_{i=1}^{T} p(\Sigma_2^{i:T}) \prod_{i=1}^{T} p(y_1^{i:T}) \prod_{i=1}^{T} p(y_2^{i:T})}
\]

(16)

For the space limit, we do not give a completely detailed result here. Based on the above posterior function, we derive the distribution density of every item conditioned with the rest and conduct Gibbs sampling. After the parameters have been estimated, predictions can be made. The conditional posterior density of each element is specified in Appendix.

The structure of the Gibbs sampler on the estimation of parameters and hidden variables is as follows:
(1) Initialize parameters $A^{(1)}$, $d^{(1)}$, $A^{(2)}$, $d^{(2)}$ and latent variables $\{\Sigma_t^{(1)}\}$, $\{\Sigma_t^{(2)}\}$.
(2) Sample $\Sigma_t^{(1)}$ from $\Sigma_t^{(1)}|\{\Sigma_{t-1}^{(1)}, A^{(1)}, d^{(1)}, A^{(2)}, d^{(2)}\}$, for $t = 1, \ldots, T$.
(3) Sequentially, sample each parameter or latent matrix from its conditional posterior distribution density. Before every drawing within one cycle, we update the known samples when producing draws for other elements.
(4) Go to Step (2) until predefined interations.

Cycling through steps (2)-(6) is a complete flow path of MCMC sampler for the coupled Wishart process. Actually, in each step of sampling, other techniques like Metropolis sampling is also applied [20], as most of the posterior densities are not the common distributions that we can directly make samples from.

VII. EXPERIMENTS AND EVALUATION

Several experiments are performed on synthetic and real-life data sets to show the effectiveness of our proposed coupled Wishart process. The experiments for this research are designed into two stages. The first part is the simulation conducted on two sets of coupled synthetic data to test the capacity of our proposed model in capturing the coupling relationship. In the second stage, we implement this model and corresponding learning methods to a real-life data set.

A. Evaluation Measures

In the following experiments, three evaluation measures, mean absolute percentage error (MAPE) [3], mean squared error (MSE) [21], and determinant error (Det_error) [3] are included to assess the quality of our method.

1) Mean Absolute Percentage Error: To evaluate the prediction quality, we use the MAPE between the predicted and true covariance matrices. The MAPE for the $(i, j)$ element is calculated as

$$MAPE_{ij} = \frac{1}{T} \sum_{t=1}^{T} \left| \frac{\Sigma_t^{-1} - E(\Sigma_t^{-1})}{\Sigma_t^{-1}} \right| \quad (17)$$

where $\Sigma_t^{-1}$ is the inverse of covariance matrix at time $t$. $E(\Sigma_t^{-1})$ is the estimated inverse covariance matrix mean, defined as $E(\Sigma_t^{-1}) = (\bar{A}^{1/2})(\Sigma_t^{-1})^t(\bar{A}^{1/2})^t$.

2) Mean Squared Error: Another evaluation metric is the MSE, usually adopted to measure the fitted time series values in statistics. With simulated data, we can directly calculate the MSE as the true value is known, see Equation (18). Here, we do not use the inverse of covariance matrices, but themselves.

$$MSE_{ij} = \frac{1}{T} \sum_{t=1}^{T} (\Sigma_t(i, j) - \Sigma_t(i, j))^2 \quad (18)$$

where $\Sigma_t$ is the estimated covariance matrix mean.

However, in real applications, we never observe the true covariance matrix $\Sigma_t$. When the ground truth is not known, we alternatively use the proxy $\tilde{S}_t(i, j) = y_t(i)y_t(j)$ where $y_t$ is the $t$th component of the multivariate observation $y_t$. This is because $E[y_t(i)y_t(j)] = \Sigma_t(i, j)$, assuming $y(t)$ has a zero mean, and a brief proof is provided in Equation (19).

$$\Sigma_t(i, j) = \text{Cov}(y_t(i), y_t(j)) = E[y_t(i) - E(y_t(i))][y_t(j) - E(y_t(j))] = E(y_t(i)y_t(j)) \quad (19)$$

In a thorough empirical study, Brownlees et al. [22] use the univariate analogue of this proxy. The MSE error then can be calculated by

$$MSE_{ij} = \frac{1}{T} \sum_{t=1}^{T} (\tilde{S}_t(i, j) - y_t(i)y_t(j))^2 \quad (20)$$

3) Determinant Error: Besides, we compute the logarithmic determinants of the estimated covariance matrix and the true value, and calculate the Det_error as

$$\text{Det_error} = \frac{1}{T} \sum_{t=1}^{T} (\log |\Sigma_t| - \log |E(\Sigma_t)|)^2 \quad (21)$$

where $\Sigma(t)$ is the estimated covariance matrix mean.

For all these three measures, the smaller ones indicate closer estimated values with the ground truth, corresponding to better models.

B. Synthetic Data Analysis

In the first part, we set all the parameters as follows to generate two time series $\{Y_t^{(1)}\}$ and $\{Y_t^{(2)}\}$:

- Process 1: $\nu^{(1)} = 30$, $d^{(1)} = 0.7$, $A^{(1)} = (130 \ 50 \ 50 \ 130)$, and $\Sigma_0^{(1)} = (3.08 \ 0.83 \ 0.83 \ 0.83)$.
- Process 2: $\nu^{(2)} = 20$, $d^{(2)} = 0.3$, $A^{(2)} = (13 \ 5 \ 5 \ 13)$, and $\Sigma_0^{(2)} = (3.08 \ 0.83 \ 0.83 \ 0.83)$.

The coupling setup is: $\alpha = 0.5$, $w_1 = 0.7$ and $w_2 = 0.3$, which means the coupling relationship in this simulation is not symmetric.

Based on the above settings, we produce two sets of the coupled time series: $\{Y_t^{(1)}\}$ and $\{Y_t^{(2)}\}$. In order to estimate these parameters, the MCMC simulation is conducted with 100000 iterations. The first 20000 draws are discarded and the remains are kept. By averaging the selected samples, we obtain the estimation of both parameters and latent covariance matrices. Then we conduct this group of experiments for ten times and analyze the statistical property of the results.

Firstly, we present the estimation of parameters in box plots. As shown in Fig. 4 and Fig. 5, some of the parameters are well estimated like $\nu^{(1)}$ and $d^{(2)}$ (i.e., the estimated values for $d^{(1)}$ and $d^{(2)}$ are 0.68 and 0.23 respectively, while the true values are 0.7 and 0.3). However, some other parameters, like $\nu^{(2)}$, $\nu^{(1)}$, and $\nu^{(2)}$ are underestimated due to its sensitivity to the appropriation of covariance matrix [3]. Just like the learning of original Wishart process in [6], some parameters are estimated with bias. As our ultimate goal is to model the covariance matrices, not the parameters, now we want to know whether this underestimation would affect the learning for covariance matrices.
mostly fit the ground truth (i.e. true values in blue) better than the uncoupled values (in green), indicating that our proposed coupled Wishart process simulates the truth more effectively. The above conclusion is also supported by Table I, in which the bold refers to better results under the three evaluation metrics: MAPE, MSE and Det_error. In Table I, we compute the MAPE and MSE of every element of the two-dimension covariance matrices from both chains, and Det_errors of each chain. All these evaluation values demonstrate that our proposed model outperforms the single Wishart process.

C. Real-life Data Analysis

For the experiments on real data, we collect the monthly returns of two industries: manufacturing (Manu for short) and consuming (Consm for short) from US stock market and Hong Kong stock market. The return series include 132 months, from January 2002 to December 2012. The descriptive statistics are provided in Table II. Our research goal is to find the volatility of these two asset portfolios, considering the interaction between these markets under the circumstance of globalization.

Before implementation with our proposed model, we first prefilter the data with an AR(1) model [5], which makes the data consistent with the model assumption: $y_t|\Sigma_t \sim \mathcal{N}_k(0, \Sigma_t)$, or more specifically, $E(y_t) = 0$. After such a preprocess, we implement the coupled Wishart process on these two data sets, under the linear coupling setup in Section V. After 20000 times of sampling, we discard the first 2000
samples, and record the rest. By averaging them, we get the estimation of all parameters and hidden covariance matrices. Just like the synthetic data, we concern the hidden covariance matrices most. After averaging the selected samples, we compare the three evaluation metrics from both the coupled Wishart process and single Wishart process, see Fig. 8, Fig. 9, and Fig. 10. Note that for the real-life data, as we do not know the true covariance matrix, a proxy proposed in Equation (20) is used in the calculation of MSE, while for the synthetic data, we refer to Equation (18).

We can see that under such three evaluation metrics, the results of coupled Wishart process are always better than those of the single Wishart process.

In summary, in the first stage of experiments, we simulate two sets of coupled time series and then conduct the experiments, results show that our proposed learning method can properly capture the coupling relationship and learn the hidden covariance matrices. Based on the conclusion obtained from stage 1, we continue to implement our proposed model on the real-life data. With comparisons, the coupled Wishart process is again demonstrated to outperform the single Wishart process.

VIII. CONCLUSION

Wishart process has been proposed to model the volatility as one effective approach with great flexibility and capturing power. In this paper, we propose a coupled volatility analysis model to capture the interaction between different systems. Based on several synchronous Wishart processes, a linear coupled Wishart process is put forward. At time \( t \), the covariance matrices exert their influences according to a learnt weight. After the model has been set up, based on the observations, we can deduce the posterior distribution of both parameters and latent variables. After that, we develop an algorithm to learn the parameters and latent covariance matrices, mainly with Gibbs sampling and and Metropolis-Hasting sampling. Experiments on synthetic data and real data show that our proposed coupled Wishart process performs better than the single Wishart process.

Although our model can effectively capture the coupling relationship when modeling the volatility of several synchronous markets, the complexity of models and simulation methods put high demand on the computation ability. When the dimension of observations or the number of Wishart processes grows, we need to re-deduce the posterior distribution and the computation cost highly increases. There
is great demand for the simplification of models and more efficient learning methods. In the future, we will improve our proposed model on such directions. In addition, this coupled model is currently used on financial data, we can also apply it to other fields, such as neurological science [17].

APPENDIX

Under our model setup, the posterior densities of all the latent covariance matrices and parameters are deduced as follows. For the limit of space, we only present the equations for chain 1. As chain 1 and chain 2 are homogeneous, the equations for chain 2 are similar.

At the periods $t = 1, 2, \ldots, T - 1$, the conditional posterior for $(\Sigma_t^{-1}, \nu_t)$ is proportional to the products below:

$$
p((\Sigma_t^{-1}, \nu_t) | \cdot) \propto \text{Wish}((\Sigma_t^{-1}), \nu_t, S_t^{-1}),
$$

$$
\text{Wish}((\Sigma_{t+1}^{-1}), \nu_t, S_t^{-1}),
$$

where $\text{Wish}((\Sigma_t^{-1}), \nu_t, S_t^{-1})$ is:

$$
\forall k/\nu \times \exp\left(-\frac{1}{2} \text{tr}(S_{t-1}^{-1} \nu_t^{-1} \Sigma_t^{-1} \nu_t^{-1} S_{t-1}^{-1})\right),
$$

$$
|\Sigma_t^{-1}|^{-1/2} \exp\left(-\frac{1}{2} (y_t^1 - y_{t-1}^1)^T \nu_t^{-1} (y_t^1 - y_{t-1}^1)\right)
$$

(22)

When $t = T$, we have:

$$
p((\Sigma_T^{-1}, \nu_T) | \cdot) \propto \text{Wish}((\Sigma_T^{-1}), \nu_T, S_T^{-1}),
$$

$$
N(y_T^1 | \Sigma_T^{-1}),
$$

where $\Sigma_T^{-1}$ is:

$$
\forall k/\nu \times \exp\left(-\frac{1}{2} \text{tr}(S_T^{-1} \nu_T^{-1} \Sigma_T^{-1} \nu_T^{-1} S_T^{-1})\right),
$$

$$
|\Sigma_T^{-1}|^{-1/2} \exp\left(-\frac{1}{2} (y_T^1 - y_{T-1}^1)^T \nu_T^{-1} (y_T^1 - y_{T-1}^1)\right)
$$

(23)

The conditional posterior distribution for $d$ is:

$$
p(d^1 | \cdot) \propto \prod_{t=1}^{T} \left[ (\Sigma_{t-1}^{-1})^{-1/2} \exp\left(-\frac{1}{2} \text{tr}(S_{t-1}^{-1} \nu_t^{-1} \Sigma_{t-1}^{-1})\right) \right]
$$

$$
\times \prod_{t=1}^{T} \left[ (\Sigma_{t-1}^{-1})^{-1/2} \exp\left(-\frac{1}{2} (y_t^1 - y_{t-1}^1)^T \nu_t^{-1} (y_t^1 - y_{t-1}^1)\right) \right]
$$

(24)

For $\nu$, the conditional posterior distribution is:

$$
p(\nu_t | \cdot) \propto p(\nu_t) \prod_{t=1}^{T} \text{Wish}(\nu_t, S_t^{-1})
$$

$$
\times \left(\frac{|\nu_t|^1 A(\nu_t^{-1})^{-1/2}}{2^{(1)/2}}\right)^T \left[ (\Sigma_{t-1}^{-1})^{(1)} \nu_t^{-1} d_t^{(1)}, \nu_t^{-1} d_t^{(1)} \right]
$$

$$
\times \exp\left(-\frac{1}{2} \sum_{t=1}^{T} \text{tr}(S_{t-1}^{-1} \nu_t^{-1} \Sigma_{t-1}^{-1})\right)
$$

(25)

REFERENCES