Forecasting Co-Volatilities via Factor Models with Asymmetry and Long Memory in Realized Covariance*

Manabu Asai

Faculty of Economics
Soka University, Japan

Michael McAleer

Department of Quantitative Finance
National Tsing Hua University, Taiwan
and
Econometric Institute
Erasmus School of Economics
Erasmus University Rotterdam
and
Tinbergen Institute
The Netherlands
and
Department of Quantitative Economics
Complutense University of Madrid, Spain

March 2014

*The authors are most grateful to Yoshi Baba for very helpful comments and suggestions. The first author acknowledges the financial support of the Japan Ministry of Education, Culture, Sports, Science and Technology, Japan Society for the Promotion of Science, and Australian Academy of Science. The second author is most grateful for the financial support of the Australian Research Council, National Science Council, Taiwan, and the Japan Society for the Promotion of Science. Address for correspondence: Faculty of Economics, Soka University, 1-236 Tangi-cho, Hachioji, Tokyo 192-8577, Japan. Email address: m-asai@soka.ac.jp.
Abstract

Modelling covariance structures is known to suffer from the curse of dimensionality. In order to avoid this problem for forecasting, the authors propose a new factor multivariate stochastic volatility (fMSV) model for realized covariance measures that accommodates asymmetry and long memory. Using the basic structure of the fMSV model, the authors extend the dynamic correlation MSV model, the conditional/stochastic Wishart autoregressive models, the matrix-exponential MSV model, and the Cholesky MSV model. Empirical results for 7 financial asset returns for US stock returns indicate that the new fMSV models outperform existing dynamic conditional correlation models for forecasting future covariances. Among the new fMSV models, the Cholesky MSV model with long memory and asymmetry shows stable and better forecasting performance for one-day, five-day and ten-day horizons in the periods before, during and after the global financial crisis.

Keywords: Dimension reduction; Factor Model; Multivariate Stochastic Volatility; Leverage Effects; Long Memory; Realized Volatility.

JEL classifications: C32, C53, C58, G17
1 Introduction

Modeling and forecasting covariance matrices of financial asset returns has been attracting interests especially for obtaining optimal portfolio of assets. Two major issues are to guarantee the positive definiteness of covariance matrices, and to reduce the number of parameters which grows with the order of squares of the number of assets. There are three classes in the fields, namely the multivariate autoregressive conditional heteroskedasticity (ARCH) class, the multivariate stochastic volatility (MSV) models, and the models of realized covariance matrix.

Regarding the multivariate ARCH family, the dynamic conditional correlation (DCC) model of Engle (2002) is most popular, and its variants are still developed in the literature. Engle (2002) suggested a structure for time-varying correlation matrices, working with a univariate ARCH class model for each conditional variance. As the estimation of the conditional variance model is conducted separately, it concentrates on the estimation of the dynamic correlations. By its structure, the DCC model reduces the number of parameters. Also with simple restrictions, Engle (2002) showed that it can guarantee the positive definiteness of the correlation matrix. Apart from the DCC model, it is worth noting the diagonal ARCH model of Ding and Engle (2001), as it enables us to estimate the model based on the estimates of unconditional covariance matrix, which can be obtained by the average of the outer-products of mean-subtracted return vectors. The approach is known as the ‘variance targeting’, and it largely decreases the number of parameters for estimating in a single step. See the survey papers of McAleer (2005), Bauwens, Laurent, and Rombouts (2006) and Silvennoinen and Teräsvirta (2009) for the models of univariate and multivariate ARCH family.

In the class of the MSV models, volatilities (variances) of stock return vector are assumed to be unobservable. Harvey, Ruiz and Shephard (1990) suggested basic MSV models with the
constant correlation matrix. As the volatilities are latent variables, it is hard to obtain the likelihood function analytically, and hence there are several techniques for estimating SV models, including the Markov chain Monte Carlo method for Jacquire, Polson and Rossi (1994) and Chib, Nardari and Shephard (2006), the efficient method of moments of Gallant and Tauchen (1996), and the Monte Carlo likelihood method proposed by Durbin and Koopman (1997), Sandmann and Koopman (1998) and Asai and McAleer (2006). See the survey papers for Asai, McAleer and Yu (2006) and Chib, Omori and Asai (2009) for alternative MSV models.

Recent developments on estimating realized co-volatility (covariance) measure of asset returns enable us to obtain consistent estimators for the unobservable volatility and co-volatilities, which can be used for estimating univariate/multivariate conditional/stochastic volatility models. In univariate SV case, Barndorff-Nielsen and Shephard (2002) and Bollerslev and Zhou (2002) suggested to use realized volatilities for estimating SV models, as the realized volatility can be considered as the true volatility with the microstructure noise. See Gourieroux Jasiak, and Sufana (2009), Chiriac and Voev (2011), Bauer and Vorkink (2011), Tao et al. (2011) and Golosnoy, Gribisch, and Liesenfeld (2012) for such applications in multivariate context.


Long memory and asymmetric effects are important features for analysis on financial time series. Regarding the long memory, Ray and Tsay (2000) introduced long range dependence into
the volatility factor model by Harvey, Ruiz, and Shephard (1994), although these papers assume the constant correlation matrix. Without any factor specifications, Chiriac and Voev (2011) suggested a long memory model, using the Cholesky decomposition of realized covariance matrices. Our factor models will involve long memory in time-varying covariances/correlations. It is worth noting that the heterogeneous autoregressive (HAR) model of Corsi (2009) and its variants are often used for approximating longer memory processes than the first order autoregressive term. See Corsi and Renò (2010), Martens, van Dijk, and de Pooter (2009), Bollerslev, Sizova, and Tauchen (2011), Chiriac and Voev (2011) and Golosnoy, Gribisch, and Liesenfeld (2012). We will also consider such HAR terms.

Turning to asymmetric effects, there are several papers that developed the approaches to incorporate negative correlation between stock returns and future volatility in the multivariate volatility models. Kroner and Ng (1998) and Kawakatsu (2006) suggested asymmetric multivariate GARCH models, while Cappiello, Engle and Sheppard (2006), Asai (2013) and Asai and So (2014) proposed asymmetric dynamic correlation models. For MSV models, Chan, Kohn, and Kirby (2006) and Asai and McAleer (2006, 2009b), Chib, Omori and Asai (2009), So and Choi (2009), Ishihara and Omori (2012), Ishihara, Omori and Asai (2013) incorporated asymmetric effects. However, there have been still no previous works on the multivariate volatility models which accommodate all of factor specifications, dynamic correlations, and asymmetric effects. The current paper will fulfill the gap.

In this paper, we will consider alternative factor MSV (fMSV) models with long memory and asymmetric effects, using realized volatilities and co-volatilities. Starting from a conventional factor model, we derive an alternative representation which decomposes the covariance matrix into the matrices of common factors and idiosyncratic errors orthogonally. Based on the decomposition,
we consider several specifications of the factor covariance matrix based on the dynamic correlation model of Asai and McAleer (2009a), the conditional autoregressive Wishart model of Golosnoy, Gribisch, and Liesenfeld (2012), the Wishart autoregressive model of Gourieroux, Jasiak and Sufana (2009), the matrix-exponential MSV model of Ishihara, Omori and Asai (2013), and the Cholesky MSV model of Chiriac and Voev (2011). We should note that Chiriac and Voev (2011) and Golosnoy, Gribisch, and Liesenfeld (2012) used the HAR term, and Chiriac and Voev (2011) considered a long memory model for their specification. In addition to these works, we extend all models by incorporating asymmetric effects and long memory (or HAR terms).

The remainder of the paper is organized as follows. Section 2 derives a new representation of the fMSV model, and discusses the specifications of asymmetric effects and long memory. Section 2 also shows the estimation procedure to obtain consistent estimators of common factors and their covariance matrices, based on the alternative form. Section 3 suggests alternative specifications for the fMSV models, and Section 4 gives an empirical example for seven stocks traded at the New York Stock Exchange.

In the following, for any positive definite matrix $X$ and real number $\alpha$, we define $X^\alpha$ by the spectral decomposition. For example, $X^{1/2}X^{1/2} = X$.

### 2 Asymmetry and Long Memory for Factor MSV Models

#### 2.1 Factor Multivariate SV Models

Let $y_t$ be $m \times 1$ vector of asset returns at time $t$. Consider the factor model

$$y_t = Bf_t + u_t,$$

where $f_t$ is a $k \times 1$ vector of common factors, $B$ is the $m \times k$ matrix of factor loading, and $u_t$ is the vector of idiosyncratic errors, which are uncorrelated with $f_t$. We can write $E(u_t|f_t, \mathcal{F}_{t-1}) = 0$. 


and $\text{Var}(\mathbf{u}_t|\mathbf{f}_t, \mathcal{F}_{t-1}) = \Sigma_u$, where $\mathcal{F}_t$ is the information set up to time $t$. We assume that the vector of common factors has the multivariate SV as

$$
\mathbf{f}_t = (\Omega_t^\dagger)^{1/2} \varepsilon_t, \quad \varepsilon_t \sim IID(\mathbf{0}, I_k),
$$

(2)

where $\Omega_t^\dagger$ is the $k \times k$ positive definite matrix. Hence, $\Omega_t^\dagger = \text{Var}(\mathbf{f}_t|\Omega_t^\dagger)$. Conditional on $\Omega_t^\dagger$, the $m \times m$ covariance matrix of $\mathbf{y}_t$ is given by

$$
\Sigma_t = B\Omega_t^\dagger B' + \Sigma_u.
$$

(3)

We will call the model (1)-(3) as the ‘factor MSV’ (fMSV) model.

For various fMSV models, Pitt and Shephard (1999), Liesenfeld and Richard (2003), and Chib, Nardari and Shephard (2006) among others restricted the factor loading matrix $B$ such that $b_{ij} = 0$ for $(i < j, i \leq k)$ and $b_{ii} = 1$ ($i \leq k$) for identification purpose. We consider an alternative specification.

There is no loss in generality in assuming $B$ in (1) satisfying the condition $B'B = I_k$. Then there exists an $m \times (m-k)$ matrix $A$ for which $A'B = O$ and $A'A = I_{m-k}$. In other words, $S = [B \ A]$ is an $m \times m$ orthogonal matrix, and $SS' = SS' = I_m$. Then we have an alternative representation of the fMSV model as

$$
\mathbf{y}_t = B\varphi_t + A\zeta_t,
$$

(4)

$$
\Sigma_t = B\Omega_t B' + A\Sigma_{\zeta\zeta} A',
$$

(5)

where $\varphi_t = \mathbf{f}_t + B'u_t$, $\zeta_t = A'u_t$, $\Omega_t = \Omega_t^\dagger + B'\Sigma_u B$ and $\Sigma_{\zeta\zeta} = A'\Sigma_u A$. We should note that

$$
E\left[\begin{pmatrix} \varphi_t \\ \zeta_t \end{pmatrix} \bigg| \Omega_t \right] = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad \text{Var}\left[\begin{pmatrix} \varphi_t \\ \zeta_t \end{pmatrix} \bigg| \Omega_t \right] = \begin{pmatrix} \Omega_t & \Sigma_{\varphi\zeta} \\ \Sigma_{\zeta\varphi} & \Sigma_{\zeta\zeta} \end{pmatrix},
$$

where $\Sigma_{\varphi\zeta} = B'\Sigma_u A$. Since $A'B = O$, the equations (4) and (5) can be considered as the ‘orthogonal decomposition’ of the fMSV model (1)-(3).
There are two merits in using the orthogonal decomposition. The first one is that we can define leverage effect from return vector to future volatilities and co-volatilities by considering a relationship between $\varphi_t$ and future values of $\Omega_t$. We will consider the asymmetric effects in the next subsection. The second merit is that we can obtain consistent estimators of $\varphi_t$ and $\Omega_t$ by the approach which will be explained later. For these reasons, we will consider modeling, estimation and forecasting of $\{\Omega_t\}$ rather than $\{\Omega_t^\dagger\}$ for the purpose of forecasting future values of $\{\Sigma_t\}$.

2.2 Asymmetric Effects on Factor Covariance

We specify asymmetric effects from the current returns, $y_t$, to the one-step-ahead volatilities and co-volatilities, $\Sigma_{t+1}$, via the covariance of the common factors, $\Omega_{t+1}$. For this purpose, we work with $k$-vector,

$$z_t = (B'\Sigma_t B)^{-1/2}(B'y_t) = \Omega_t^{-1/2} \varphi_t = (\Omega_t^\dagger + B'\Sigma u B)^{-1/2}(f_t + B'u_t).$$

By definition, we have $E(z_t|\Omega_t) = 0$ and $\text{Var}(z_t|\Omega_t) = I_k$, and thus $E(z_t|\Omega_t) = 0$ and $\text{Var}(z_t|\Omega_t) = I_k$.

The two popular approaches for introducing asymmetry for volatility dynamics are employing threshold function, as in the GJR model of Glosten, Jagannathan and Runkle (1992), and working with an absolute value function, as in the specification of Nelson (1991). Kroner and Ng (1998) developed a multivariate extension of the GJR model, while Kawakatsu (2006) suggested an extension for the EGARCH model. Unlike these approaches, we consider the second order Hermit polynomial for the error term. Hansen, Huang and Shek (2012) suggested this approach for a univariate model. In this case, we have

$$\lambda_{it} = \lambda_{1i} z_{it} + \lambda_{2i} (z_{it}^2 - E(z_{it}^2)).$$

Noting that $E(z_{it}^2) = 1$, we also have $E(\lambda_{it}) = 0$, $\text{Var}(\lambda_{it}) = \lambda_{1i}^2 + 2\lambda_{2i}^2$, and $E[z_{it}\lambda_{it}] = \lambda_{1i}$, as
we assume that each element of $z_t$ has zero mean and unite variance. Hence, if $\lambda_{1i}$ is negative, the model describes the leverage effect. On the other hand, $\lambda_{2i}$ is expected to be positive, as in the conventional GARCH families. We will use the vector specification, $\lambda_t = (\lambda_{1t}, \ldots, \lambda_{kt})'$, especially when we consider asymmetric effects on the diagonal elements of $\Omega_{t+1}$.

Now we suggest the following multivariate extension of the above approach:

$$\Lambda_t = \Lambda_1 \circ (z_{it}k_t + \iota_k z_i') + \Lambda_2 \circ (z_t z_t' - E(z_t z_t')) .$$

(8)

Noting that $E(z_t z_t') = I_k$, it is straightforward to show that $E(\Lambda_t) = O$ and that each element has finite variance. For the leverage effects, $\lambda_{ii,1} < 0$ and $\lambda_{ii,2} > 0$. For the purpose of considering a positive semi-definite matrix of the multivariate asymmetric effects, we modify (8) as

$$\Lambda_t^\dagger = \Lambda^\dagger \circ (z_t - \gamma)(z_t - \gamma)'$$

(9)

where $\gamma$ is $k$-vector of parameters with $\gamma_i > 0$, and $\Lambda^\dagger$ is a $k$ dimensional positive (semi-)definite matrix.

We may accommodate the asymmetric effects to the fMSV models, by introducing the error vector, $\lambda_t$, the error matrix, $\Lambda_t$ and/or the positive semi-definite matrix, $\Lambda_t^\dagger$. We will develop the new fMSV models in Section 3.

We can examine news impacts from an asset return to its future volatility, following the works of Engle and Ng (1993), Yu (2005), Asai and McAleer (2009b), Caporin and McAleer (2011) and Chen and Ghysels (2010). Engle and Ng (1993) developed the news impact curve (NIC), which is a useful tool for measuring the effects of news on the conditional variances. They showed, graphically, the asymmetric reactions of the conditional variances to positive and negative shocks of equal magnitude for the GJR model of Glosten et al. (1992) and the EGARCH model of Nelson (1991). Regarding multivariate conditional volatility models, Caporin and McAleer (2011)
developed news impact surfaces specifically for their dynamic asymmetric multivariate GARCH. In the framework of univariate SV models, Yu (2005) developed the news impact function for evaluating the effects of news on the log-volatilities as an adaptation of the NIC. Asai and McAleer (2009b) extended it to apply volatilities and log-volatilities for MSV with leverage models. When realized volatility is available, we can estimate the NIC using the MIDAS (mixed data sampling) regressions, as suggested by Chen and Ghysels (2010).

Since we accommodate factors which reduce the number of parameters, we need to check whether or not the fMSV model with asymmetric effects can approximate true NICs. We may define the news impact surface for the fMSV model as

$$\Sigma_t^{NIS}(y) = E \left( \Sigma_t | y_{t-1} = y, \Omega_{t-1} = \bar{\Omega}, \Omega_{t-2} = \bar{\Omega}, \ldots \right)$$

$$= B \left\{ E \left( \Omega_t | y_{t-1} = y, \Omega_{t-1} = \bar{\Omega}, \Omega_{t-2} = \bar{\Omega}, \ldots \right) \right\} B' + A \Sigma \zeta \zeta A'$$

where $y$ is an $m$-vector and $\bar{\Omega} = E(\Omega_t)$. By definition, we can examine news impact from one asset return not only to its own future volatility byt also to future volatilities of remaining assets. We will give illustrative examples in empirical analysis.

2.3 Long Memory for Covariance structure


Let $L$ be the lag operator. For any $k$-dimensional symmetric matrix process $X_t$, we define a
$k$-dimensional matrix $D(L)$ as

$$
D(L) = \begin{pmatrix}
(1 - L)^{d_{11}} & (1 - L)^{d_{12}} & \cdots & (1 - L)^{d_{1k}} \\
(1 - L)^{d_{21}} & (1 - L)^{d_{22}} & \cdots & (1 - L)^{d_{2k}} \\
\vdots & \vdots & \ddots & \vdots \\
(1 - L)^{d_{k1}} & (1 - L)^{d_{k2}} & \cdots & (1 - L)^{d_{kk}}
\end{pmatrix},
$$

(10)

where $d_{ij} = d_{ji}$ and $0 \leq d_{ij} \leq 1$, in order to describe the fractional difference of $X_t$ as $D(L) \circ X_t$.

We do not consider negative values of $d_{ij}$, as the empirical evidences of the fractional integration are found in the positive region in financial time series. For any process of $k$-vector $x_t$, we also define a $k$-dimensional diagonal matrix $D_k(L)$ as

$$
D_k(L) = \text{diag}((1 - L)^{d^v_1}, \ldots, (1 - L)^{d^v_k}),
$$

(11)

where $0 \leq d^v_i \leq 1$, so that $D_k(L)x_t$ is the fractional difference of $x_t$. We can separately estimate the order of the fractional difference, $d_{i,j}$ and/or $d^v_i$, via the multivariate Gaussian semiparametric estimation suggested by Shimotsu (2007), which is considered as a multivariate extension of the local Whittle (LW) estimator of Shimotsu and Phillips (2006).

Recent papers including Andersen, Bollerslev and Meddahi (2011), Bollerslev, Sizova and Tauchen (2011), Bollerslev and Todorov (2011), Chiriac and Voev (2011) work with the heterogeneous autoregressive (HAR) model of Corsi (2009) and its variants in order to describe longer-range dependence than that of AR(1) models. Following Corsi (2009), define $h$ days average of past values of $X_t$ and $x_t$ as

$$
X_{t-1,t-h} = \frac{1}{h}(X_{t-1} + \cdots + X_{t-h}), \quad x_{t-1,t-h} = \frac{1}{h}(x_{t-1} + \cdots + x_{t-h}),
$$

(12)

respectively. In addition to daily effect, $X_{t-1}$, we can consider weekly and monthly effects as $X_{t-1,t-5}$ and $X_{t-1,t-22}$, respectively.
2.4 Estimation of Factor Covariance

Consider a consistent estimator of $\Sigma_t$, which is a realized covariance matrix, and we denote it as $\hat{\Sigma}_t$. By using the sequence of realized covariance matrix, Tao et al. (2011) developed a technique to estimate the factor covariance matrix, $\Omega_t$. We will shortly explain the approach. Define

$$\bar{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \Sigma_t, \quad C = \frac{1}{T} \sum_{t=1}^{T} (\Sigma_t - \bar{\Sigma})^2.$$

Recalling that $A'B = O$, we obtain $A'\Sigma_t = A'\Sigma_u$ from equation (3). Hence, we obtain that $A'\Sigma_t = A'\bar{\Sigma}$ and

$$A'CA = \frac{1}{T} \sum_{t=1}^{T} (A'\Sigma_t - A'\bar{\Sigma}) (\Sigma_t A - \bar{\Sigma}A) = O$$

From this result, we can consider $A$ as the $m - k$ orthonormal eigenvectors of $C$, corresponding to the $(m - k)$-fold eigenvalue 0. The remaining $k$ orthonormal eigenvectors of $C$, corresponding to the $k$ nonzero eigenvalues, can be taken as the columns of the factor matrix $B$.

Regarding the consistent estimator, $\hat{\Sigma}_t$, let

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^{T} \hat{\Sigma}_t, \quad \hat{C} = \frac{1}{T} \sum_{t=1}^{T} (\hat{\Sigma}_t - \hat{\bar{\Sigma}})^2.$$ \hfill (13)

Then we obtain the estimator $\hat{B}$ using the $k$ orthonormal eigenvectors of $\hat{C}$, corresponding to the $k$ largest eigenvalues, as its columns. Consequently, we obtain the estimated factor and its estimated covariance matrix as

$$\hat{\varphi}_t = \hat{B}'\hat{y}_t, \quad \hat{\Omega}_t = \hat{B}'\hat{\Sigma}_t\hat{B}, \quad t = 1, \ldots, T,$$

respectively. We can also estimate $\Sigma_{\zeta\zeta}$ by $\hat{\Sigma}_{\zeta\zeta} = \hat{A}'\hat{\Sigma}\hat{A}$. Tao et al. (2011) showed the consistency of $\hat{B}$ and $\hat{\Omega}_t$ for fixed $k$, under several kinds of realized covariance matrices developed by Barndorff-Nielsen et al. (2008, 2011), Christensen, Kinnebrock, and Podolskij (2010), Griffin and
Oomen (2011), Hautsch, Kyj, and Oomen (2012), Wong and Zou (2010) and Zhang (2011). It is straightforward to show the consistency of $\hat{\varphi}_t$ and $\hat{\Sigma}_\zeta \zeta$ by applying the framework of Tao et al. (2011).

Once we obtained the estimates of the common factors, $\hat{\varphi}_t$, and covariance matrix of common factors, $\hat{\Omega}_t$, we can estimate alternative models for the covariance matrix of the common factors.

3 Specifications of Factor Covariance

We will extend five kinds of models for the covariance matrix, in order to accommodating asymmetry and long memory on the factor covariance matrix, $\Omega_t$.

3.1 Dynamic Correlation fMSV Models

As in Engle (2002), we start from the decomposition of the covariance matrix,

$$\Omega_t = V_t^{1/2} P_t V_t^{1/2}, \quad (14)$$

where $V_t$ is the diagonal matrix of volatilities of the vector of common factors, defined by $V_t = \text{diag}(v_{1t}, \ldots, v_{kt})$, and $P_t$ is its correlation matrix. By definition, we can write $V_t = (I_k \circ \Omega_t)$, where ‘$\circ$’ denotes the Hadamard (element-by-element) product. As in Engle (2002), we consider a positive definite matrix, $Q_t$, in order to describe the correlation dynamics with $P_t = Q_t^{s-1/2} Q_t^* Q_t^{s-1/2}$, where $Q_t^* = (I_k \circ Q_t)$. Following most of the works on MSV models, we consider the vector of log-volatilities, $h_t = (h_{1t}, \ldots, h_{kt})' = (\log v_{1t}, \ldots, \log v_{kt})'$, rather than volatilities themselves.

We use the dynamic correlation MSV (DC MSV) model of Asai and McAleer (2009a), in order
to specify the process of \((Q_t, h_t)\) as

\[
Q_t^{-1} \sim W(R_t, \delta), \quad R_t = \frac{1}{\delta} \left( Q_{t-1}^{-1} \right)^{\frac{\alpha}{2}} \bar{R} \left( Q_{t-1}^{-1} \right)^{\frac{\alpha}{2}},
\]

(15)

\[
h_t = \mu + \Phi_1 \{ h_{t-1} - \mu \} + \eta_t, \quad \eta_t \sim N(0, \Sigma),
\]

(16)

where \(\delta (\delta > k - 1)\) and \(\alpha (|\alpha| < 1)\) are scalars, \(\bar{R}\) and \(\Sigma\) are \(k\) dimensional positive definite matrices, \(\mu\) is \(k\)-vector of parameters, \(\Phi_1\) is \(k\) dimensional diagonal matrix defined by \(\Phi = \text{diag}(\phi_1, \ldots, \phi_k)\), and \(W(S, p)\) denotes the Wishart distribution with the scale matrix \(S\) and the degrees-of-freedom parameter \(p\). Here, we define \((P_{t-1})^{\alpha/2}\) by the spectral decomposition. We impose restrictions that the diagonal elements of \(\bar{R}\) are ones, and that \(Q_t = \bar{V} R_t \bar{V}^T\), where \(\bar{V} = E(V_t)\).

Now we introduce the following general model for \(Q_t\), accommodating asymmetry and long memory.

\[
Q_t^{-1} \sim W(R_t, \delta), \quad R_t = \delta \left( Q_{t-1}^{-1} \right)^{\frac{\alpha_m}{2}} \left( P_{t-1,t-5}^{-1} \right)^{\frac{\alpha_w}{2}} \left( P_{t-1,t-22}^{-1} \right)^{\frac{\alpha_m}{2}} \times \left[ w \bar{R} + (1 - w) \Lambda_t \right] \left( P_{t-1,t-5}^{-1} \right)^{\frac{\alpha_w}{2}} \left( P_{t-1,t-22}^{-1} \right)^{\frac{\alpha_m}{2}},
\]

(17)

where \(\alpha_d, \alpha_w\) and \(\alpha_m\) are scalar parameters, \(w (0 < w \leq 1)\) is a scalar weight parameter, \(P_{t-1,t-5}\) and \(P_{t-1,t-22}\) are weekly and monthly averages defined by (12), and \(\Lambda_t\) is the positive semi-definite matrix defined by equation (9) such that the \((i,j)\)th elements of \(\Lambda_t^+\) is \(\frac{1}{(1+\gamma_{ij})(1+\gamma_{ij})}\). While \(P_{t-1,t-5}\) and \(P_{t-1,t-22}\) produce longer memory than the single \(P_{t-1}\), \(\Lambda_t\) yields asymmetric effects on the correlation dynamics.

Regarding the vector of log-volatilities, \(h_t\), a general model is given by

\[
\Phi(L)D_k(L)\{ h_t - \mu \} = \lambda_{t-1} + \eta_t,
\]

(18)
where \( D_k(L) \) is the fractional difference operator defined by equation (11), \( \lambda_t \) is vector of asymmetric function defined by equation (7). We set \( \Phi(L) = I_k - \Phi_1 L \). The competing heterogeneous VAR specification is given by

\[
h_t = \Phi_d \{ h_{t-1} - \mu \} + \Phi_w \{ h_{t-1,t-5} - \mu \} + \Phi_m \{ h_{t-1,t-22} - \mu \} + \lambda_{t-1} + \eta_t, \tag{19}
\]

where \( \Phi_d, \Phi_w \) and \( \Phi_m \) are diagonal matrices.

It is convenient to have acronyms for the various fMSV-DC models: (i) \textit{fMSV-DC} means the basic model defined by equations (14)-(16); (ii) \textit{fMSV-DC-cAvA} stands for the fMSV-DC model with asymmetric effects on correlation and volatility dynamics, defined by equations (14), (17) and (18) such that \( \alpha_w = \alpha_m = 0 \) and \( D_k(L) = I_k \); (iii) \textit{fMSV-DC-cHvL} represents the fMSV-DC model with heterogeneous-time effects on correlation dynamics and with long-memory in volatility, defined by equations (14), (17) and (18) such that \( w = 1, \Lambda_t^{\dagger} = O \) and \( \lambda_t = O \); (iv) \textit{fMSV-DC-cHvH} is the fMSV-DC model with heterogeneous-time effects on correlation and volatility dynamics, defined by equations (14), (17) and (19) such that \( w = 1, \Lambda_t^{\dagger} = O \) and \( \lambda_t = O \); (v) \textit{fMSV-DC-cAHvAL} means the fMSV-DC model with asymmetric and heterogeneous-time effects on correlation dynamics and with asymmetric effects and long-memory in volatility dynamics, defined by equations (14), (17) and (18); (vi) \textit{fMSV-DC-cAHvAH} stands for the fMSV-DC model with asymmetric and heterogeneous-time effects on correlation and volatility dynamics, defined by equations (14), (17) and (19).

We can estimate the fMSV-DC model (14), (17) and (18) (or (19)) using the likelihood function based on the Wishart distribution, when \( \Omega_t \) is available. As noted in previous section, we can estimate long memory parameter \( d_i^\star \) separately. In order to reduce the number of parameters in estimation, we may estimate the diagonal elements of \( \bar{V} \) by the diagonal elements of the average of \( \bar{\Omega}_t \).
3.2 Wishart Disturbance Models for fMSV

Philipov and Glickman (2006a, b), Golosnoy, Gribisch and Liesenfeld (2012) and Asai and So (2013) assume that the covariance matrix follows a Wishart distribution (or an inverse Wishart distribution) with parameters that depend on the past covariance matrix. For the case of Wishart distribution, we can introduce a disturbance matrix for the covariance matrix of \( \varphi_t \) as

\[
\Omega_t = H_t^{1/2} \Xi_t H_t^{1/2}, \quad \Xi_t \sim W\left((1/\nu)I_k, \nu\right),
\]

where \( H_t \) is a \( k \) dimensional positive definite matrix, determined by the information up to \( t-1 \). By the property of the Wishart distribution, \( E(\Xi_t) = I_k \) and thus \( E(\Omega_t|\Omega_{t-1}, \Omega_{t-1}^\varphi) = H_t \), where \( \Omega_{t-1}^\varphi \) consists of \( \{(\varphi_{t-1}, \Omega_{t-1}), (\varphi_{t-2}, \Omega_{t-2}), \ldots\} \). By definition, we also have \( \Omega_t | H_t \sim W\left((1/\nu)H_t, \nu\right) \).

Analogous to the BEKK model of Engle and Kroner (1995) and the diagonal MGARCH model of Ding and Engle (2001), we consider the following two specifications

\[
H_t = W + G^\prime \Omega_{t-1} G + K^\prime H_{t-1} K,
\]

\[
H_t = (\iota^\prime - G^* - K^*) \circ \Omega^* + G^* \circ \Omega_{t-1} + K^* \circ H_{t-1},
\]

where \( W \) and \( \Omega^* \) are \( k \) dimensional positive definite matrices of parameters, \( G \) and \( K \) are \( k \times k \) matrices, \( \iota \) is a \( k \)-vector of ones, and \( F^* \) and \( G^* \) are \( k \) dimensional positive (semi-)definite matrices.

By putting \( E(\Omega_t) = E(H_t) \) in equation (22), we have \( \Omega^* = E(\Omega_t) \), which is the unconditional expectation of \( \Omega_t \). The equations (21) and (22) have the same parsimonous specification, which is obtained by setting \( G = \text{diag}(g) \), \( K = \text{diag}(k) \) and \( W = (\iota^\prime - gg^\prime - kk^\prime) \circ \Omega^* \) in equation (21), namely,

\[
H_t = (\iota^\prime - gg^\prime - kk^\prime) \circ \Omega^* + (gg^\prime) \circ \Omega_{t-1} + (kk^\prime) \circ H_{t-1},
\]

where \( g \) and \( k \) are \( k \)-vector of parameters.
We will work with the Wishart disturbance model (20) and (23) in our analysis. We will call the model as the ‘conditional autoregressive Wishart fMSV’ (fMSV-CAW) model, following Golosnoy, Gribisch and Liesenfeld (2012).

The general specification for the fMSV-CAW model is given by equation (20) and

\[
H_t = (\eta^t - g_d g_d^t - g_w g_w^t - g_m g_m^t - k k^t) \circ \Omega^* + (k k^t) \circ H_{t-1}
\]

\[
+ (g_d g_d^t) \circ \Omega_{t-1} + (g_w g_w^t) \circ \Omega_{t-1,t-5} + (g_m g_m^t) \circ \Omega_{t-1,t-22} + \Lambda_t \Gamma_{t-1}^\dagger - \bar{\Lambda}^\dagger,
\]

where \(\Omega_{t-1,t-5}\) and \(\Omega_{t-1,t-22}\) are weekly and monthly averages defined by (12), \(\Lambda_t^\dagger\) is asymmetric matrix function defined by (9), \(\bar{\Lambda}^\dagger = E(\Lambda_t^\dagger) = \Lambda^\dagger \circ (I_k + \gamma \gamma')\) and \(\Lambda^\dagger = \lambda^\dagger \lambda^\dagger'\) with \(k\)-vector \(\lambda^\dagger\).

The acronyms for the various fMSV-CAW models is as follows: (i) fMSV-CAW means the basic fMSV-CAW model, defined by equations (20) and (23); (ii) fMSV-CAW-A means the fMSV-CAW model with asymmetric effects, defined by equations (20) and (24) such that \(g_w = g_m = 0\); (iii) fMSV-CAW-H means the fMSV-CAW model with heterogeneous-time effects, defined by equations (20) and (24) such that \(\Lambda_t^\dagger = \bar{\Lambda}^\dagger = O\) and \(k = 0\); (iv) fMSV-CAW-AH means the fMSV-CAW model with asymmetric and heterogeneous-time effects, defined by equations (20) and (24) with \(k = 0\).

When \(\Omega_t\) is available, we can estimate these models by minimizing the log-likelihood function based on the Wishart distribution.

3.3 Wishart Autoregressive fMSV Models

Gourieroux, Jasiak and Sufana (2009) employed non-central Wishart distribution to define autoregressive property of a process of covariance matrices. They defined the model by the conditional
moment generating function, which is

\[ E[\exp(\text{tr}\{Y\Omega_t\})|\Omega_{t-1}] = \frac{\exp(\text{tr}\{G'Y(I_k - 2\Omega^\ast Y)^{-1}G'\Omega_{t-1}\})}{[\text{det}(I_k - 2\Omega^\ast Y)]^{\nu/2}}, \]  

(25)

where \( G, W \) and \( \nu \) are the same as in equations (20) and (21), and \( Y \) is a \( k \times k \) symmetric matrix which satisfies \( \|W^{1/2}YW^{1/2}\| < 1 \). By Proposition 2 of Gourieroux, Jasiak and Sufana (2009), we can write

\[ \Omega_t = \nu W + G'\Omega_{t-1}G + N_t, \]  

(26)

where \( N_t \) is a heteroskedastic error term with conditional mean zero. The conditional mean of \( \Omega_t \) in (26) is similar to the specification in (21), but the approach of introducing error term is different.

As argued in Laurent, Rombouts, and Violante (2012) variance targeting estimation is useful when forecasting conditional covariance matrices. In this case, we may replace \( \nu W \) with \( \Omega^\ast - G'\Omega^\ast G \), where \( \Omega^\ast = E(\Omega_t) \).

The general specification of the fMSV-WAR model is given by

\[ \Omega_t = (\nu - g_dg_d' - g wg_w' - g mg_m' - kk') \circ \Omega_{t-1} + \Lambda_{t-1}^\dagger - \Lambda_{t-22}^\dagger + (g_dg_d') \circ \Omega_{t-1} + (g wg_w') \circ \Omega_{t-1,t-5} + (g mg_m') \circ \Omega_{t-1,t-22} + N_t, \]  

(27)

where parameters are the same as fMSV-CAW models. Note that the fMSV-WAR models has no past conditional covariance matrix unlike the fMSV-CAW, by construction. The acronyms for the various fMSV-WAR models is as follows: (i) \textbf{fMSV-WAR} means the basic fMSV-WAR model, defined by equation (26); (ii) \textbf{fMSV-WAR-A} means the fMSV-WAR model with asymmetric effects, defined by equation (27) such that \( g_w = g_m = 0 \); (iii) \textbf{fMSV-WAR-H} means the fMSV-WAR model with heterogeneous-time effects, defined by equation (27) such that \( \Lambda_{11}^\dagger = \Lambda_{11}^\dagger = O; \)
(iv) **fMSV-WAR-AH** means the fMSV-WAR model with asymmetric and heterogeneous-time effects, defined by equation (27).

Rather than conducting ML estimation via non-central Wishart distribution, Gourieroux, Jasiak and Sufana (2009) suggest to estimate via method of moments with first two moments. We will follow the latter approach.

### 3.4 Matrix-Exponential fMSV Models


For any square matrix \( X \), the matrix-exponential transformation is defined by \( \text{Exp}(X) = \sum_{i=0}^{\infty} \frac{1}{i!} X^i \) with \( X^0 = I \). The same result is obtained by working with the spectral decomposition, as we have \( \text{Exp}(X) \) by replacing the eigenvalues by their exponential transformation. Note that \( \text{Exp}(X) \) is positive definite, whenever \( X \) is symmetric. In the same manner, \( \text{Log}(Y) \) is defined by its spectral decomposition of a positive definite matrix, \( Y \), with replacement of the logarithmic transformation of eigenvalues.

Following Ishihara, Omori and Asai (2012), we consider the following matrix-exponential fMSV (ME-fMSV) model.

\[
\Omega_t = \text{Exp}(\Theta_t),
\]

\[
\Theta_t = M + \Psi_1 \circ \{ \Theta_{t-1} - M \} + E_t,
\]

where \( \xi_t = \text{vech}(E_t) \sim N(0, \Sigma_\xi) \), \( \Theta_t \) are \( k \times k \) symmetric matrices, \( M \) and \( \Psi_1 \) is the \( k \) dimensional
symmetric matrix of parameters, $E_t$ are $k$ dimensional symmetric matrix of disturbance, and $\Sigma_\xi$ is the $k(k + 1)/2$ dimensional positive definite matrix.

The matrix-exponential fMSV model with asymmetry and long-memory is given by equation (28) and

$$
\Psi(L) \circ D(L) \circ \{ \Theta_t - M \} = \Lambda_{t-1} + E_t,
$$

(30)

where $\Psi(L) = \nu' - \Psi_1 L$, the asymmetric error matrix $\Lambda_t$ is defined by (8), and the matrix of fractional difference operator $D(L)$ is defined by (11). Instead of long-memory, we can incorporate heterogeneous-time effect into the fMSV model as,

$$
\Theta_t = M + \Psi_d \circ \{ \Theta_{t-1} - M \} + \Psi_w \circ \{ \Theta_{t-1,t-5} - M \} + \Psi_m \circ \{ \Theta_{t-1,t-22} - M \} + \Lambda_{t-1} + E_t,
$$

(31)

where $\Psi_d$, $\Psi_w$ and $\Psi_m$ is the $k$ dimensional symmetric matrix of parameters. The acronyms for the various fMSV-EXP models is as follows: (i) fMSV-EXP means the basic fMSV-EXP model, defined by equation (29); (ii) fMSV-EXP-A means the fMSV-EXP model with asymmetric effects, defined by equation (30) such that $D(L) = \nu'$; (iii) fMSV-EXP-L means the fMSV-EXP model with long memory, defined by equation (30) such that $\Lambda_1 = \Lambda_2 = 0$; (iv) fMSV-EXP-AL means the fMSV-EXP model with asymmetry and long memory, defined by equation (30); (v) fMSV-EXP-H means the fMSV-EXP model with heterogeneous-time effects, defined by equation (31) such that $\Lambda_1 = \Lambda_2 = 0$; (vi) fMSV-EXP-AH means the fMSV-EXP model with asymmetric and heterogeneous-time effects, defined by equation (31).

When $\Omega_t$ is available, we can calculate $\Theta_t = \log(\Omega_t)$ in order to conduct the OLS estimation for each equation. As noted in previous section, we can estimate long memory parameter $d_{ij}$ separately.
3.5 Cholesky fMSV Models

Chiriac and Voev (2011) applied the Cholesky decomposition for modeling realized covariances. For the covariance matrix of $\phi_t$, we consider the Cholesky decomposition of $\Omega_t$ as

$$\Omega_t = L_t L_t', \quad (32)$$

where $L_t$ are lower triangular matrices. Define $a_t = \text{vech}(L_t)$. Then we can consider VAR model for $a_t$ as,

$$a_t = \mu_c + \Phi_c \{a_{t-1} - \mu_c\} + \xi_t \quad (33)$$

where $\mu_c$ is the $k(k+1)/2$ vector of parameters, $\Phi_c$ is the $k(k+1)/2$ dimensional diagonal matrix, and $\xi_t$ is the $k(k+1)/2$ vector of error terms.

Analogous to the fMSV-EXP models, we can develop fMSV-CH, fMSV-CH-A, fMSV-CH-L, fMSV-CH-AL, fMSV-CH-H, and fMSV-CH-AH models. When $\Omega_t$ is available, we can calculate $a_t$ in order to conduct the OLS estimation for each equation. As noted in previous section, we can estimate long memory parameter $d_{ij}$ separately.

4 Empirical Analysis

4.1 Data and Preliminary Analysis

We examine forecasting performances of five classes of fMSV models using daily realized covariance matrices for seven stocks traded at the New York Stock Exchange: Alcoa Inc. (AA), American Express (AXP), Bank of America (BAC), du Pont de Nemours and Company (DD), General Electric (GE), International Business Machines (IBM), and Coca-Cola Company (KO). Based on the vector of returns for the $m = 7$ stocks computed for 1-min interval of trading day at $t$ between 9:30 a.m. and 4:00 p.m., we calculated daily realized volatilities and co-volatilities by
the averaging approach of Wang and Zou (2010) using 5-min averages with adjustments. Note that the method of Wang and Zou (2010) is robust to microstructure noise, and we can use the estimator for estimating factor covariances as proposed by Tao et al. (2011). As suggested by Bickel and Levina (2008a,b), Johnstone and Lu (2009) and Wang and Zou (2010), we impose some sparsity structure on the estimator in order to guarantee the consistency. For this purpose these authors work with threshold which means retaining its elements whose absolute values exceed a given value and replacing others by zero. For convenience, we find the threshold for co-volatilities on each day $t$, so that the estimated covariance matrix is positive definite. We also calculated the corresponding open-close returns for seven assets. The vector of returns and its covariance matrix estimator are denoted by $y_t$ and $\hat{\Sigma}_t$, respectively. The sample period starts at February 20, 1997, and ends on October 4, 2012, covering 3900 observations. Table 1 presents the descriptive statistics of the returns, volatilities and co-volatilities. The empirical distribution of the returns is heavily skewed to the left and is highly leptokurtic. Regarding volatilities and co-volatilities, they are skewed to the right except for co-volatilities for (AXP-KO) and (BAC-DD), with evidences of heavy-tails for all series. It should be noted that extremely large values for volatilities are observed during the IT bubble in the US.

We will use latter 1500 observations for forecasting analysis later in this section. Figure 1 shows the estimated volatilities for the period of forecasting. Figure 1 indicates that volatilities are high in the period of turbulence caused by the GFC. Figure 2 presents some of estimated correlation dynamics. We chose ten series which have stronger correlations than remaining eleven series. Because of the threshold explained above, correlation coefficients often take zero when they are close to zero. An interesting feature is that the correlation dynamics of (AA-BAC), (AXP-BAC), (GE-BAC), (IBM-BAC) fluctuate around zero just after the bankruptcy of Lehman
Brothers, then they show stronger correlation on average. One year after the bankruptcy of Lehman Brothers, these four series shows weak correlations around zero. On the contrary, other correlation processes show strong connection for the whole period for forecasting.

As a preliminary analysis with former $T = 2400$ observations, Table 2 gives the contribution rates of eigenvalues of $\hat{C}$, defined by equation (13). As reported in Sugiyama and Tong (1976) and Sheena (2013), the estimator of contribution rates are suffered from bias in finite sample, and hence Table 2 also gives the modified estimator of Sheena (2013). In this case, there are no major differences in these two estimators of contribution rates. The cumulative percentage of the eigenvalues up to the first 4 components is 89 percent, while it is 95% for the 5 components. For this reason, we will employ 4 and 5 factors in our empirical analysis.

We also estimated the news impact surface based on the estimates from the fMSV-CAW-A model, using the former periods from 1997 to 2006. Figure 3 gives NICs the NIC from $y_{i,t}$ to $\sigma_{j,t+1}^{1/2}$ for $i, j = \text{AA, AXP, BAC}$, after removing the effects from other assets. For convenience, the news impact is scaled by the mean of realized volatility for each series. While the NIC from each asset to its own future volatility shows asymmetric effects, the news impacts for other volatilities are negligible. Regarding BAC, the positive relation between current return and future volatility reflects the rapid growth for the years from 2000 to 2006.

### 4.2 Benchmark Models

For the benchmark models, we employ two kinds of dynamic conditional correlation (DCC) models; one is the asymmetric DCC (ADCC) model suggested by Cappiello, Engle and Sheppard (2006), while the other is the fractionally integrated matrix-exponential DCC (FIEDCC) model of Asai and So (2014). While ADCC model is popular and it captures the asymmetric effects in dynamic correlations, the FIEDCC model accommodates the long-memory and asymmetry in the dynamic
correlation process.

We consider the conventional specification of dynamic correlation model as
\[ y_t = m_t + u_t, \quad E(u_t | 3_{t-1}) = 0, \quad \text{Var}(u_t | 3_{t-1}) = \Sigma_t = V_t^{1/2} P_t V_t^{1/2}, \]
\[ V_t = \text{diag}(v_t), \quad P_t = Q_t^{*-1/2} Q_t^{*-1/2}, \quad Q_t^{*} = (I_k \circ Q_t), \]
where \( m_t = E(y_t | 3_{t-1}) \) are the conditional mean of \( y_t \), \( v_t = (v_{it}, \ldots, v_{mt})' \) are the vector of conditional volatilities, and \( Q_t \) are \( m \) dimensional positive definite matrices which determine the process of conditional correlations. We also use the standardization defined by \( z_t = V_t^{-1/2}(y_t - m_t) \), which produces \( E(z_t z_t' | 3_{t-1}) = P_t \). We may simply assume that \( m_t = 0 \).

Our ADCC model is given by equation (34) and
\[ Q_t = (P - aP - bP - gN) + g(z_{i-1}z_{i-1}') + g(z_{i-1}z_{i-1}'), \]
\[ v_t = \omega + \delta \circ u_{t-1} \circ u_{t-1} + \eta \circ v_{t-1} + \xi \circ u_{t-1} \circ u_{t-1}, \]
where \( \omega, \delta, \eta \) and \( \xi \) are vectors of parameters, \( a, b \) and \( g \) are scalar parameters, \( z_{i-1} = (z_{it}, \ldots, z_{mt})' \), \( z_{it} = z_{it}I(z_{it} < 0), I(\cdot) \) is an indicator function which takes the value of 1 if the argument is true and 0 otherwise, \( P = E(z_t z_t') \) and \( N = E(z_{t-1} z_{t-1}') \). We work with the GJR model for the volatility process due to its status as one of popular asymmetric models in the ARCH family. As proposed by Cappiello, Engle and Sheppard (2006), we replace \( P \) and \( N \) by the sample covariance matrices.

The second model is the FIEDCC model given by equation (34) and
\[ (1 - aL)(1 - L)^d (\log Q_t - M) = \lambda_1(z_{i-1}z_{i-1} + \tau_m z_{i-1}' - P), \]
\[ (1 - \phi_i L)(1 - L)^{d_i}(\log u_{it} - \xi_i) = \lambda_i z_{i,t-1} + \lambda_2_i (z_{i,t-1}^2 - 1), \]
where \( d_i \) and \( d_i^* \) are scalar parameters of fractional difference, \( a \) and \( \phi_i \) are AR(1) coefficients, \( \lambda_1i, \lambda_2i, \lambda_1i \) and \( \lambda_2i \) are scalar parameters for error terms which accommodate asymmetric effects, \( M = E(\log Q_t), \xi_i = E(\log u_{it}) \). As suggested by Asai and So (2014), we specify \( M = \log(P) \) and replace \( P \) by the sample covariance matrix.

We employ these two DCC models as the benchmarks for the empirical analysis.
4.3 Results

We set $T = 2400$ for estimating alternative forecasting models, focusing on predict horizons of $h = \{1, 5, 10\}$ days. We denote the $h$-step ahead forecast of $\Sigma_t$ as $\hat{\Sigma}^f_{t+h}$. We re-estimate every model fixing sample size as $T = 2400$, and obtain new forecasts based on updated parameters. In our forecasting analysis, we use three out-of-sample windows, taking account for the effects of the Global Financial Crisis (GFC). The first window covers the period from September 8, 2006 to September 12, 2008. The second period starts from the bankruptcy of Lehman Brothers, i.e. September 15, 2008, and ends on September 24, 2010. The third window covers the period from September 27, 2010 to October 4, 2012. We will call these three windows as periods ‘before GFC’ (B-GFC), ‘during GFC’ (D-GFC) and ‘after GFC’ (A-GFC), respectively. The size of each window is 500.

We use $\hat{\Sigma}_{t+h}$ as a proxy for the unobservable $\Sigma_{t+h}$, and define forecast error matrix as

$$E_{t+h} = \hat{\Sigma}_{t+h} - \hat{\Sigma}^f_{t+h},$$

where $\hat{\Sigma}^f_{t+h}$ is the $h$-step-ahead forecast of $\Sigma_t$. Following Chiriac and Voev (2011) and Golosnoy, Gribisch, and Liesenfeld (2012), we compare the seven models’ out-of-sample forecast root-mean-squared error (RMSE) based on the Frobenius norm of the forecast error, which is defined by

$$FN_h = \frac{1}{T_h} \sum_t ||E_{t+h}|| = \frac{1}{T_h} \sum_t \left[ \sum_{i,j} e_{ij,t+h}^2 \right]^{1/2},$$

(37)

where $T_h$ is the number of forecast periods. Especially, $T_1 = 500$, $T_5 = 496$ and $T_{10} = 491$.

Table 3 shows the results for forecasting performances. For 5- and 10-step-ahead forecasts, the statistic is standardized by adjusting by the number of steps to make the results comparable. Among the benchmark DCC models, the FIEDCC model always performs better than the ADCC model. However, all the factor MSV models performs better than the FIEDCC model.
Before examining the results of out-of-sample forecasts, we compare the results of in-sample forecasts for $T = 2400$ observations, presented by Table 3. For the 4 factor MSV models, the DC-cAH-vAL, DC-cAH-vAH, EXP, EX-A, EXP-H and EXP-AH have smaller RMSE than the other models. Regarding the 5 factor models, the DC-cAH-vAL, DC-cAH-vAH, EXP-L, EX-AL, EXP-H and EXP-AH perform better than the other models. The WAR type and CAW type models give relatively larger RMSEs.

Regarding the out-of-sample forecasts, Table 3 implies three general results. First of all, the results obviously depend on the time periods, ‘before’, ‘during’ and ‘after’ the GFC, rather than the length of forecasting horizons. Compare to the RMSEs before the GFC, those for the period during the GFC is about 1.2 times, while those for the period after GFC is less than the half. The sizes of volatilities in Figure 1 support the results. Secondly, Table 3 indicate that there are no major differences between 4 and 5 factor models, showing the success of approximating via factors. Thirdly, accommodating asymmetric effects generally improves the RMSE of 1-day-ahead forecasts. We may improve the results by introducing asymmetric effects for heterogeneous time horizons, as in Chen and Ghysels (2010) and Asai (2013).

The best forecasting model depends on the time period and forecasting horizons. However, the CH-L and CH-AL models have relatively small RMSEs for all cases, supporting the results of Chiriac and Voev (2011). Although the simple WAR and WAR-A shows the best performance during the GFC, the CH-L and CH-AL models competitive results. In addition to these models, the DC-cAH-vAL, EXP-L and EXP-AL models often give the competitive results. We should note the similarity of the structure of CAW-type and WAR-type models, and differences of the forecasting performances brought by the differences of estimation procedures. Minimizing the difference of first moment in WAR-type models has an advantage on the above $FN_h$. 

26
5 Conclusion

In this paper, we developed an approach for modeling asymmetry and long memory on the factor MSV (fMSV) model, using its alternative representation. Based on the approach, we extended DC, CAW, WAR, matrix-exponential MSV and Cholesky MSV model for specifying the covariance matrix of common factors. In forecasting application with seven stocks traded at the New York Stock Exchange, we show that the new fMSV models perform better than the asymmetric DCC and fractionally-integrated matrix-exponential DCC models. Among the new fMSV models, the Cholesky fMSV model with long memory and asymmetry is the best model, since it has smallest RMSEs (or competitive results), and since it is robust to forecasting horizons and the periods before, during and after the GFC.

Our new factor approach for asymmetric effects and long memory opens to many interesting research directions. We may include not only daily asymmetric effects but also weekly and monthly effects, as in Chen and Ghysels (2010) and Asai (2013). We neglected the jumps and the impact of macroeconomic variables on future volatility and co-volatility. For these issues, we need to wait further researches on co-jumps and parsimonious specifications for macroeconomic variables.
References


Table 1: Descriptive Statistics for the Returns, Realized Volatilities and Co-Volatilities

<table>
<thead>
<tr>
<th>Stock</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>Std.Dev.</th>
<th>Skew.</th>
<th>Kurt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>-0.1290</td>
<td>-70.939</td>
<td>13.737</td>
<td>2.8079</td>
<td>-8.0234</td>
<td>200.61</td>
</tr>
<tr>
<td>AXP</td>
<td>0.0270</td>
<td>-107.15</td>
<td>14.458</td>
<td>2.8477</td>
<td>-13.455</td>
<td>511.57</td>
</tr>
<tr>
<td>BAC</td>
<td>-0.0871</td>
<td>-67.659</td>
<td>20.080</td>
<td>2.9418</td>
<td>-3.3396</td>
<td>82.829</td>
</tr>
<tr>
<td>DD</td>
<td>-0.0201</td>
<td>-67.922</td>
<td>10.052</td>
<td>2.1052</td>
<td>-8.5250</td>
<td>277.95</td>
</tr>
<tr>
<td>GE</td>
<td>0.0145</td>
<td>-71.076</td>
<td>14.125</td>
<td>2.2669</td>
<td>-14.702</td>
<td>453.93</td>
</tr>
<tr>
<td>IBM</td>
<td>-0.0376</td>
<td>-66.247</td>
<td>11.748</td>
<td>2.0435</td>
<td>-8.6308</td>
<td>283.87</td>
</tr>
<tr>
<td>KO</td>
<td>0.0288</td>
<td>-11.158</td>
<td>9.9392</td>
<td>1.4375</td>
<td>8.6970</td>
<td>8.6970</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stock</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>Std.Dev.</th>
<th>Skew.</th>
<th>Kurt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>6.1297</td>
<td>0.0561</td>
<td>3797.4</td>
<td>74.397</td>
<td>41.868</td>
<td>1920.3</td>
</tr>
<tr>
<td>AXP</td>
<td>6.4466</td>
<td>0.0610</td>
<td>9624.3</td>
<td>153.43</td>
<td>62.408</td>
<td>3912.4</td>
</tr>
<tr>
<td>BAC</td>
<td>5.7171</td>
<td>0.0813</td>
<td>2847.6</td>
<td>47.653</td>
<td>54.040</td>
<td>3206.1</td>
</tr>
<tr>
<td>DD</td>
<td>3.7965</td>
<td>0.0118</td>
<td>3101.2</td>
<td>52.793</td>
<td>53.705</td>
<td>3056.3</td>
</tr>
<tr>
<td>GE</td>
<td>3.1839</td>
<td>0.0612</td>
<td>2713.5</td>
<td>46.938</td>
<td>51.792</td>
<td>2882.3</td>
</tr>
<tr>
<td>IBM</td>
<td>3.4917</td>
<td>0.0857</td>
<td>2680.2</td>
<td>43.241</td>
<td>51.792</td>
<td>3720.5</td>
</tr>
<tr>
<td>KO</td>
<td>1.7498</td>
<td>0.0174</td>
<td>605.39</td>
<td>9.9454</td>
<td>56.786</td>
<td>3438.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stock</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>Std.Dev.</th>
<th>Skew.</th>
<th>Kurt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA</td>
<td>1.4739</td>
<td>-59.937</td>
<td>137.88</td>
<td>4.3539</td>
<td>11.621</td>
<td>299.80</td>
</tr>
<tr>
<td>AXP</td>
<td>1.0984</td>
<td>-42.190</td>
<td>84.610</td>
<td>3.7854</td>
<td>7.6459</td>
<td>112.70</td>
</tr>
<tr>
<td>BAC</td>
<td>0.4821</td>
<td>-30.114</td>
<td>26.276</td>
<td>1.2324</td>
<td>0.5513</td>
<td>159.21</td>
</tr>
<tr>
<td>DD</td>
<td>0.8989</td>
<td>-3.6298</td>
<td>76.535</td>
<td>2.2981</td>
<td>14.243</td>
<td>350.98</td>
</tr>
<tr>
<td>GE</td>
<td>1.3141</td>
<td>-2.8196</td>
<td>125.46</td>
<td>3.5966</td>
<td>15.833</td>
<td>435.35</td>
</tr>
<tr>
<td>IBM</td>
<td>0.3413</td>
<td>-11.475</td>
<td>16.942</td>
<td>0.9482</td>
<td>4.9720</td>
<td>82.635</td>
</tr>
<tr>
<td>KO</td>
<td>1.7498</td>
<td>0.0174</td>
<td>605.39</td>
<td>9.9454</td>
<td>56.786</td>
<td>3438.9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Stock</th>
<th>Mean</th>
<th>Min</th>
<th>Max</th>
<th>Std.Dev.</th>
<th>Skew.</th>
<th>Kurt.</th>
</tr>
</thead>
<tbody>
<tr>
<td>AA-AXP</td>
<td>1.4739</td>
<td>-59.937</td>
<td>137.88</td>
<td>4.3539</td>
<td>11.621</td>
<td>299.80</td>
</tr>
<tr>
<td>AA-BAC</td>
<td>1.0984</td>
<td>-42.190</td>
<td>84.610</td>
<td>3.7854</td>
<td>7.6459</td>
<td>112.70</td>
</tr>
<tr>
<td>AA-DD</td>
<td>0.4821</td>
<td>-30.114</td>
<td>26.276</td>
<td>1.2324</td>
<td>0.5513</td>
<td>159.21</td>
</tr>
<tr>
<td>AA-GE</td>
<td>0.8989</td>
<td>-3.6298</td>
<td>76.535</td>
<td>2.2981</td>
<td>14.243</td>
<td>350.98</td>
</tr>
<tr>
<td>AA-IBM</td>
<td>1.3141</td>
<td>-2.8196</td>
<td>125.46</td>
<td>3.5966</td>
<td>15.833</td>
<td>435.35</td>
</tr>
<tr>
<td>AA-KO</td>
<td>0.3413</td>
<td>-11.475</td>
<td>16.942</td>
<td>0.9482</td>
<td>4.9720</td>
<td>82.635</td>
</tr>
<tr>
<td>AXP-BAC</td>
<td>1.6783</td>
<td>-14.746</td>
<td>81.759</td>
<td>5.1727</td>
<td>7.6358</td>
<td>81.302</td>
</tr>
<tr>
<td>AXP-DD</td>
<td>0.5802</td>
<td>-27.327</td>
<td>59.436</td>
<td>1.7817</td>
<td>11.081</td>
<td>356.76</td>
</tr>
<tr>
<td>AXP-GE</td>
<td>1.0623</td>
<td>-11.736</td>
<td>44.476</td>
<td>2.1931</td>
<td>7.5522</td>
<td>99.438</td>
</tr>
<tr>
<td>AXP-IBM</td>
<td>1.5056</td>
<td>-34.963</td>
<td>99.649</td>
<td>3.6069</td>
<td>9.8920</td>
<td>193.75</td>
</tr>
<tr>
<td>AXP-KO</td>
<td>0.4554</td>
<td>-134.08</td>
<td>19.622</td>
<td>2.5052</td>
<td>-38.759</td>
<td>2114.3</td>
</tr>
<tr>
<td>BAC-DD</td>
<td>0.5001</td>
<td>-65.812</td>
<td>19.828</td>
<td>1.7622</td>
<td>-10.627</td>
<td>532.2</td>
</tr>
<tr>
<td>BAC-GE</td>
<td>0.8557</td>
<td>-15.578</td>
<td>50.629</td>
<td>2.1967</td>
<td>7.9632</td>
<td>117.02</td>
</tr>
<tr>
<td>BAC-IBM</td>
<td>1.3033</td>
<td>-7.1569</td>
<td>96.732</td>
<td>4.4039</td>
<td>10.500</td>
<td>157.94</td>
</tr>
<tr>
<td>BAC-KO</td>
<td>0.4435</td>
<td>-7.9434</td>
<td>15.860</td>
<td>1.1695</td>
<td>5.4442</td>
<td>57.057</td>
</tr>
<tr>
<td>DD-GE</td>
<td>0.5024</td>
<td>-5.2749</td>
<td>107.63</td>
<td>2.0553</td>
<td>37.012</td>
<td>1878.1</td>
</tr>
<tr>
<td>DD-IBM</td>
<td>0.569</td>
<td>-7.0968</td>
<td>21.231</td>
<td>1.3087</td>
<td>5.5390</td>
<td>58.218</td>
</tr>
<tr>
<td>DD-KO</td>
<td>0.4330</td>
<td>-7.0070</td>
<td>18.929</td>
<td>1.0510</td>
<td>6.1137</td>
<td>74.406</td>
</tr>
<tr>
<td>GE-KO</td>
<td>0.4258</td>
<td>-8.6030</td>
<td>16.380</td>
<td>1.0004</td>
<td>5.6546</td>
<td>62.400</td>
</tr>
<tr>
<td>IBM-KO</td>
<td>0.5188</td>
<td>-24.579</td>
<td>18.491</td>
<td>1.2088</td>
<td>2.7633</td>
<td>96.704</td>
</tr>
</tbody>
</table>

Note: The number of observations for each series is 3900.
Table 2: Estimated Contribution Rates of Eigenvalues of $C$

<table>
<thead>
<tr>
<th>Component</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimates</td>
<td>0.6182</td>
<td>0.1440</td>
<td>0.0737</td>
<td>0.0578</td>
<td>0.0550</td>
<td>0.0486</td>
<td>0.0027</td>
</tr>
<tr>
<td>Modified Estimates</td>
<td>0.6172</td>
<td>0.1438</td>
<td>0.0737</td>
<td>0.0578</td>
<td>0.0550</td>
<td>0.0486</td>
<td>0.0027</td>
</tr>
</tbody>
</table>

Note: We use the approach of Sheena (2013) to calculate modified estimators of contribution rates.
Table 3: RMSE based on the Frobenius Norm of the Forecasting Error

<table>
<thead>
<tr>
<th>Model</th>
<th>In-sample</th>
<th>1 day-ahead forecasts</th>
<th>5 day-ahead forecasts</th>
<th>10 day-ahead forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B-GFC</td>
<td>D-GFC</td>
<td>A-GFC</td>
<td>B-GFC</td>
</tr>
<tr>
<td>ADCC</td>
<td>4.8929</td>
<td>4.3007</td>
<td>7.6422</td>
<td>1.9062</td>
</tr>
<tr>
<td>Benchmark Models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 factor MSV models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DC</td>
<td>2.6840</td>
<td>3.7689</td>
<td>4.6917</td>
<td>1.6633</td>
</tr>
<tr>
<td>DC-cHvL</td>
<td>2.3811</td>
<td>3.7885</td>
<td>5.0152</td>
<td>1.6520</td>
</tr>
<tr>
<td>DC-cHvH</td>
<td>2.2003</td>
<td>3.8116</td>
<td>5.5722</td>
<td>1.7002</td>
</tr>
<tr>
<td>CAW-A</td>
<td>2.8897</td>
<td>3.8953</td>
<td>5.1027</td>
<td>1.6814</td>
</tr>
<tr>
<td>CAW-H</td>
<td>2.9988</td>
<td>3.9266</td>
<td>6.3134</td>
<td>1.7590</td>
</tr>
<tr>
<td>CAW-AH</td>
<td>2.9929</td>
<td>3.9322</td>
<td>5.4074</td>
<td>1.7491</td>
</tr>
<tr>
<td>WAR</td>
<td>2.5504</td>
<td>3.9842</td>
<td>4.5689</td>
<td>1.6137</td>
</tr>
<tr>
<td>WAR-A</td>
<td>2.5502</td>
<td>3.9846</td>
<td>4.5642</td>
<td>1.6135</td>
</tr>
<tr>
<td>EXP</td>
<td>2.1941</td>
<td>3.7901</td>
<td>5.0084</td>
<td>1.6602</td>
</tr>
<tr>
<td>EXP-H</td>
<td>2.1948</td>
<td>3.7988</td>
<td>5.3729</td>
<td>1.6896</td>
</tr>
<tr>
<td>EXP-AH</td>
<td>2.1939</td>
<td>3.8003</td>
<td>5.2919</td>
<td>1.6882</td>
</tr>
<tr>
<td>CH</td>
<td>2.2600</td>
<td>3.7792</td>
<td>5.1872</td>
<td>1.6528</td>
</tr>
<tr>
<td>CH-A</td>
<td>2.2600</td>
<td>3.7572</td>
<td>5.0136</td>
<td>1.6175</td>
</tr>
<tr>
<td>CH-AL</td>
<td>2.2511</td>
<td>3.7342</td>
<td>4.6603</td>
<td>1.6008</td>
</tr>
<tr>
<td>CH-H</td>
<td>2.2188</td>
<td>3.7745</td>
<td>5.5917</td>
<td>1.6848</td>
</tr>
<tr>
<td>CH-AH</td>
<td>2.2187</td>
<td>3.7763</td>
<td>5.5588</td>
<td>1.6851</td>
</tr>
</tbody>
</table>

Note: ‘B-GFC’, ‘D-GFC’ and ‘A-GFC’ stand for ‘Before GFC’, ‘During GFC’ and ‘After GFC’. For 5- and 10-step-ahead forecasts, the statistic is standardized by adjusting by the number of steps to make the results comparable.
<table>
<thead>
<tr>
<th>Model</th>
<th>In-sample</th>
<th>1 day-ahead forecasts</th>
<th>5 day-ahead forecasts</th>
<th>10 day-ahead forecasts</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>B-GFC</td>
<td>D-GFC</td>
<td>A-GFC</td>
<td>B-GFC</td>
</tr>
<tr>
<td>5 factor MSV models</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>DC</td>
<td>2.7771</td>
<td>3.8026</td>
<td>4.7625</td>
<td>1.7066</td>
</tr>
<tr>
<td>DC-cAhA</td>
<td>2.3503</td>
<td>3.8421</td>
<td>5.6042</td>
<td>1.7177</td>
</tr>
<tr>
<td>DC-cHvL</td>
<td>2.3612</td>
<td>3.8072</td>
<td>5.0389</td>
<td>1.6962</td>
</tr>
<tr>
<td>DC-cHvH</td>
<td>2.1373</td>
<td>3.8318</td>
<td>5.8305</td>
<td>1.7365</td>
</tr>
<tr>
<td>DC-cAhvAH</td>
<td>2.1302</td>
<td>3.8261</td>
<td>5.7057</td>
<td>1.7301</td>
</tr>
<tr>
<td>CAW</td>
<td>2.9490</td>
<td>3.8642</td>
<td>5.0261</td>
<td>1.9058</td>
</tr>
<tr>
<td>CAW-AH</td>
<td>2.9477</td>
<td>3.9112</td>
<td>5.8848</td>
<td>1.6662</td>
</tr>
<tr>
<td>EXP</td>
<td>2.3227</td>
<td>3.7363</td>
<td>4.6218</td>
<td>1.6317</td>
</tr>
<tr>
<td>EXP-A</td>
<td>2.3120</td>
<td>3.7340</td>
<td>4.6220</td>
<td>1.6288</td>
</tr>
<tr>
<td>EXP-L</td>
<td>2.1247</td>
<td>3.7925</td>
<td>4.9899</td>
<td>1.6714</td>
</tr>
<tr>
<td>EXP-AL</td>
<td>2.1238</td>
<td>3.7695</td>
<td>4.8031</td>
<td>1.6402</td>
</tr>
<tr>
<td>EXP-H</td>
<td>2.1183</td>
<td>3.8017</td>
<td>5.4224</td>
<td>1.7080</td>
</tr>
<tr>
<td>EXP-AH</td>
<td>2.1174</td>
<td>3.8036</td>
<td>5.3638</td>
<td>1.7077</td>
</tr>
<tr>
<td>CH</td>
<td>2.2674</td>
<td>3.7916</td>
<td>5.1831</td>
<td>1.6827</td>
</tr>
<tr>
<td>CH-A</td>
<td>2.2672</td>
<td>3.7645</td>
<td>4.9832</td>
<td>1.6470</td>
</tr>
<tr>
<td>CH-H</td>
<td>2.1438</td>
<td>3.7827</td>
<td>5.6585</td>
<td>1.7020</td>
</tr>
<tr>
<td>CH-AH</td>
<td>2.1439</td>
<td>3.7820</td>
<td>5.6414</td>
<td>1.7026</td>
</tr>
</tbody>
</table>

Note: ‘B-GFC’, ‘D-GFC’ and ‘A-GFC’ stand for ‘Before GFC’, ‘During GFC’ and ‘After GFC’. For 5- and 10-step-ahead forecasts, the statistic is standardized by adjusting by the number of steps to make the results comparable.
Figure 1: Time Series Plots of Realized Volatilities

Note: The periods covers September 8, 2006 to October 4, 2012, giving 1,500 observations.
Figure 2: Time Series Plots of Realized Correlations

Note: The periods covers September 8, 2006 to October 4, 2012, giving 1,500 observations.
Figure 3: News Impact Curves from Returns to Future Volatilities

Note: Figure shows the news impact curves from $y_{i,t}$ to $\sigma_{j,t+1}^{1/2}$ ($i, j = AA, AXP, BAC$).