The generalized conditional autoregressive Wishart model for multivariate realized volatility

Philip L.H. Yu, W.K. Li & F.C. Ng


To link to this article: http://dx.doi.org/10.1080/07350015.2015.1096788

Accepted author version posted online: 20 Oct 2015.

Submit your article to this journal

Article views: 43

View Crossmark data
The generalized conditional autoregressive Wishart model for multivariate realized volatility

Philip L.H. Yu, W.K. Li and F.C. Ng
Department of Statistics and Actuarial Science
The University of Hong Kong
Pokfulam Road, Hong Kong, China

Abstract

It is well known that in finance variances and covariances of asset returns move together over time. Recently, much interest has been aroused by an approach involving the use of the realized covariance (RCOV) matrix constructed from high-frequency returns as the ex-post realization of the covariance matrix of low-frequency returns. For the analysis of dynamics of RCOV matrices, we propose the generalized conditional autoregressive Wishart (GCAW) model. Both the noncentrality matrix and scale matrix of the Wishart distribution are driven by the lagged values of RCOV matrices, and represent two different sources of dynamics, respectively. The GCAW is a generalization of the existing models, and accounts for symmetry and positive definiteness of RCOV matrices without imposing any parametric restriction. Some important properties such as conditional moments, unconditional moments and stationarity are discussed. Empirical examples including sequences of daily RCOV matrices from the New York Stock Exchange illustrate that our model outperforms the existing models in terms of model fitting and forecasting.
1 Introduction

Multivariate volatility modeling plays a critical role in the areas of risk management, portfolio optimization and asset pricing. It is widely believed that variances and covariances of asset returns move together over time. Therefore, understanding and forecasting the temporal and cross-sectional dependence in elements of the covariance matrix are of vital importance in financial econometrics. The most popular approach should be multivariate GARCH models (reviewed in Bauwens et al., 2006) which treats the conditional covariance matrix of returns as a deterministic function of past returns. The other common approach is multivariate stochastic volatility models (reviewed in Asai et al., 2006) in which the volatility process is assumed to be random. The common feature is that the covariance matrix is regarded as a latent variable and must be modeled together with the underlying returns.

Thanks to the availability of high-frequency return data, much interest has been aroused in recent years by an alternative approach that involves direct modeling of the covariance matrix. Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004) showed that the realized covariance (RCOV) matrix constructed from high-frequency return data is a precise estimate for the covariance matrix of low-frequency returns. Since then, there is a huge literature on constructing RCOV matrix estimator, including Aït-Sahalia et al. (2010), Barndorff-Nielsen et al. (2011), Christensen et al. (2010), Fan et al. (2015), Shephard and Xiu (2014), Wang and Zou (2010), and Zhang (2011), to name only a few. As such, in this approach, the daily covariance matrix is first estimated by the RCOV matrix, making it effectively observable. Then, the observed RCOV matrices can be modeled directly. For example, Tao et al. (2011) fits a vector autoregressive model to the estimated volatility factors which are obtained from fitting a matrix factor model to the observed RCOV matrices.

An important requirement of multivariate volatility modeling is that each covariance matrix in the process at any time point must be symmetric and positive definite. Some methods employ appropriate transformations of the covariance matrix to take into account of this requirement. Bauer and Vorkink (2011) transforms the RCOV matrix by the matrix logarithm function and then assumes heterogeneous autoregressive processes (Corsi, 2009) for the transformed volatilities. Chiriac and Voev (2011) decomposes the RCOV matrix by the Cholesky decomposition and then models the Cholesky factors with fractionally integrated processes. One drawback of this approach is that direct interpretation of parameters becomes difficult since the models are applied.
to the transformed series instead of the RCOV matrix.

Another approach to cope with symmetry and positive definiteness is to make use of matrix-variate distributions with support restricted to the set of all symmetric and positive definite matrices. A natural choice is the Wishart distribution. A noncentral Wishart distribution involves three parameters: the degrees of freedom, noncentrality matrix and scale matrix. The central Wishart distribution is a special case with zero noncentrality matrix.

Asai and So (2010) and Jin and Maheu (2013) considered the joint modeling of returns and covariances based on the central Wishart distribution. The former model is based on the normal-inverse-Wishart transition distribution where the conditional inverse covariance matrix follows a central Wishart distribution and the scale matrix has a dynamic correlation specification. The latter model assumes that the conditional covariance matrix follows a central Wishart distribution with the scale matrix driven by sample averages of lagged RCOV matrices.

Two pioneer time-series models for multivariate RCOVs based on the Wishart distribution are found in Gouriéroux et al. (2009) and Golosnoy et al. (2012). Gouriéroux et al. (2009) proposed the Wishart autoregressive (WAR) model which involves a Wishart transition distribution with a time-varying noncentrality matrix and a constant but nonzero scale matrix. It is well known that the sampling distribution of the sample covariance matrix estimator for independent and identically distributed multivariate normal random vectors is the central Wishart distribution (Wishart, 1928). The WAR model extends it by assuming an autoregressive multivariate normal random vectors. Many studies in the literature of RCOV (e.g. Hansen and Lunde, 2006; Oomen, 2006; Bandi and Russell, 2008; Aït-Sahalia et al., 2011) pointed out that the RCOV matrix is biased due to autoregressive high-frequency returns which is induced by market microstructure noise. The WAR model incorporates this when modeling the dynamics of RCOV matrices, which is contrary to the common practice of using not too frequent intra-day data to construct the RCOV matrix in order to reduce the effect of market microstructure noise. However, it assumes homoscedasticity (a constant scale matrix) which contradicts the stylized fact that asset returns have a time-varying (conditional) covariance matrix.

On the other hand, Golosnoy et al. (2012) proposed the conditional autoregressive Wishart (CAW) model which involves a central Wishart transition distribution with a time-varying scale matrix that has the BEKK-GARCH specification of Engle and Kroner (1995). Although the CAW model has no clear specification of the underlying return processes as in the WAR model, one
may realize a close analogy between the CAW model for RCOV matrices and the BEKK-GARCH model for multivariate conditional heteroscedastic returns.

To summarize, both the WAR and CAW approaches model the RCOV matrix through the Wishart transition distribution with different sources of dynamics. The WAR model has a time-varying noncentrality matrix that originates from autoregressive high-frequency returns, whereas the CAW model has a time-varying scale matrix that is the result of conditional heteroscedastic returns. It is interesting to ask whether the two sources are statistically significant. If both of them are significant, incorporating them should yield a better analysis of RCOVs.

In this paper, we propose the generalized conditional autoregressive Wishart (GCAW) model, a new class of dynamic model for multivariate RCOVs based on the Wishart distribution, which generalizes both the WAR and CAW. Both the noncentrality matrix and scale matrix of the Wishart distribution are time-varying and represent two different sources of dynamics as mentioned above. As such, it accounts for complex temporal interdependence across elements of the RCOV matrix. The RCOV matrix is guaranteed to be symmetric and positive definite without imposing any parametric restriction, and the parameters can be easily estimated by the Maximum Likelihood Estimation. Some important time series properties such as conditional moments, unconditional moments and stationarity are discussed. The model is applied to sequences of daily RCOV matrices from the New York Stock Exchange. Inferences on the parameters are made to investigate if the time-varying noncentrality matrix and scale matrix are needed and which one contributes more to the dynamics of RCOV matrices. We also compare the forecasting performance of the proposed model with the existing models.

The rest of the paper is organized as follows. In Section 2, we present the GCAW model. Simulation studies are carried out in Section 3. Section 4 provides an empirical example, and we conclude in Section 5.
2 Generalized Conditional Autoregressive Wishart (GCAW) model

2.1 Model specification

Consider the stochastic, symmetric and positive definite RCOV matrix $Y_t = (Y_{ijt})$ of $n$ asset returns observed at time $t$ ($t = 1, \ldots, T$). The matrix $Y_t$ given the past observations $\mathcal{F}_{t-1} = \{Y_{t-1}, Y_{t-2}, \ldots\}$ is assumed to follow a Wishart distribution:

$$Y_t \mid \mathcal{F}_{t-1} \sim W_n(v, \Lambda_t, \Sigma_t),$$

where $v > n - 1$ is the degrees of freedom, $\Lambda_t = (\Lambda_{ijt})$ is the $n \times n$ symmetric and positive semi-definite noncentrality matrix and $\Sigma_t = (\Sigma_{ijt})$ is the $n \times n$ symmetric and positive definite scale matrix; and has a density function (see Muirhead, 1982, p. 442):

$$f(Y_t \mid \mathcal{F}_{t-1}) = 2^{-n/2} \pi^{-n(n-1)/4} \left[ \prod_{i=1}^{n} \Gamma \left( \frac{v + 1 - i}{2} \right) \right]^{-1} (\det \Sigma_t)^{-v/2} (\det Y_t)^{(v-n-1)/2} \times \exp \left\{ \frac{1}{2} \text{tr} \left[ \Sigma_t^{-1} (Y_t + \Lambda_t) \right] \right\} _0 \text{F}_1 \left( \frac{v}{2}, \frac{1}{4} \Sigma_t^{-1} \Lambda_t \Sigma_t^{-1} Y_t \right),$$

where $\Gamma(\cdot)$ is the gamma function and $_0\text{F}_1$ is the hypergeometric function of matrix argument. In order to account for temporal and cross-sectional dependence in elements of $Y_t$, we assume that $\Lambda_t$ and $\Sigma_t$ are driven by the lagged values of $Y_t$:

$$\Lambda_t = \sum_{k=1}^{r} M_k Y_{t-k} M_k' \tag{3}$$

$$\Sigma_t = C C' + \sum_{i=1}^{p} B_i \Sigma_{t-i} B_i' + \sum_{j=1}^{q} A_j Y_{t-j} A_j' \tag{4}$$

where $C$ is an $n \times n$ lower triangular matrix and $A_j$, $B_i$, $M_k$ are $n \times n$ parameter matrices. Equation (3) accounts for the autoregressive property of high-frequency returns. Equation (4) resembles the BEKK-GARCH specification of Engle and Kroner (1995) and accounts for conditional heteroscedasticity. Without imposing any parametric restriction, equation (3) guarantees a symmetric and positive semi-definite noncentrality matrix $\Lambda_t$, and equation (4) guarantees a symmetric and positive definite scale matrix $\Sigma_t$ as long as the initial matrices $\Sigma_0$, $\Sigma_{-1}, \ldots, \Sigma_{-p+1}$ are symmetric and positive definite.
Similar to other models based on the Wishart distribution, each RCOV matrix of the process is guaranteed to be symmetric and positive definite. However, the model defined by equations (1)–(4) is unidentified. To make the model identifiable, we restrict the main diagonal elements of $C$, i.e. $C_{11}, \ldots, C_{nn}$, and the first diagonal element of each $A_j, B_i, M_k$, i.e. $A_{11,1}, \ldots, A_{11,q}, B_{11,1}, \ldots, B_{11,p}, M_{11,1}, \ldots, M_{11,r}$, to be positive (see Engle and Kroner, 1995).

The GCAW model is a generalization of the W AR model of Gouriéroux et al. (2009) and the CAW model of Golosnoy et al. (2012). If $M_1 = \cdots = M_r = 0$ (i.e., $r = 0$), the GCAW($p, q, r$) model becomes the CAW($p, q$) model. If $A_1 = \cdots = A_q = B_1 = \cdots = B_p = 0$ (i.e., $p = q = 0$), the GCAW($p, q, r$) model becomes the W AR($r$) model. As such, by comparing the GCAW($p, q, r$) model with the CAW($p, q$) and W AR($r$) model respectively, we can test if the time-varying noncentrality matrix and scale matrix are needed in the analysis of RCOV matrices based on the Wishart distribution. For instance, to test if at least one of $M_k$’s $\neq 0$ ($k = 1, \ldots, r$) (i.e., the time-varying noncentrality matrix is important), we can use the likelihood ratio test with the CAW($p, q$) model as the restricted model (the null hypothesis) and the GCAW($p, q, r$) model as the unrestricted model (the alternative hypothesis).

2.2 Conditional moments

Denote vech(·) as the operator that stacks the lower triangular part including the diagonal of a matrix into a vector, and vec(·) as the operator that stacks all columns of a matrix into a vector. Let $L_n$ and $D_n$ denote the elimination and duplication matrix, respectively, such that, for any symmetric $n \times n$ matrix $Y$, vec($Y$) = $D_n$vech($Y$) and vech($Y$) = $L_n$vec($Y$) (see Turkington, 2002, p. 43-44). Define $y_t$ = vech($Y_t$), $\lambda_t$ = vech($\Lambda_t$), $\sigma_t$ = vech($\Sigma_t$) and $c$ = vech($CC'$). Then, the vector representation of equations (3) and (4) become

$$\lambda_t = \sum_{k=1}^{r} M_k y_{t-k},$$

$$\sigma_t = c + \sum_{i=1}^{p} B_i \sigma_{t-i} + \sum_{j=1}^{q} A_j y_{t-j},$$

where $M_k, A_j$ and $B_i$ are square matrices with dimension $N = n(n+1)/2$ such that

$$M_k = L_n(M_k \otimes M_k)D_n, \quad A_j = L_n(A_j \otimes A_j)D_n, \quad B_i = L_n(B_i \otimes B_i)D_n.$$
and \( \otimes \) denotes the Kronecker product.

In order to define the standardized residuals for model diagnostics which will be described in Section 2.5, the first two conditional moments of \( y_t \) are required.

**Proposition 1.** The conditional expectation and covariance matrix for the GCAW\((p, q, r)\) model defined by equations (1)–(4) are given by

\[
E(y_t | F_{t-1}) = \lambda_t + v\sigma_t = vc + \sum_{i=1}^{p} vB_i \sigma_{t-i} + \sum_{j=1}^{\max(q, r)} (vA_j + M_j)y_{t-j}, \tag{7}
\]

\[
\text{Var}(y_t | F_{t-1}) = 2D_n^+ \left[ v(\Sigma_t \otimes \Sigma_t) + \Sigma_t \otimes \Lambda_t + \Lambda_t \otimes \Sigma_t \right] (D_n^+)', \tag{8}
\]

with \( A_j = 0 \) for \( j > q \) and \( M_j = 0 \) for \( j > r \), where \( D_n^+ = (D_n' D_n)^{-1} D_n \) is the Moore–Penrose generalized inverse of \( D_n \).

A special case is the WAR\((r)\) model, where \( A_1 = \cdots = A_q = B_1 = \cdots = B_p = 0 \). By equation (7), the GCAW\((p, q, r)\) model gives a more flexible conditional expectation specification than the WAR\((r)\) model, which is also driven by the lagged values of \( \sigma_t \). Another special case is the CAW\((p, q)\) model, where \( \Lambda_t = 0 \). By equation (8), the conditional covariance matrix of \( y_t \) under the GCAW\((p, q, r)\) model has a more complex structure than that under the CAW\((p, q)\) model.

The conditional covariance matrix of \( y_t \) given in equation (8) is quite cumbersome, with a very large dimension \( \frac{n(n+1)}{4} \left[ \frac{n(n+1)}{2} + 1 \right] \). In order to understand the structure of that matrix, we derive the conditional covariance between two inner products \( \gamma' Y_t \alpha \) and \( \delta' Y_t \beta \).

**Proposition 2.** Under the GCAW\((p, q, r)\) model defined by equations (1)–(4), for any \( n \times 1 \) vectors \( \alpha, \beta, \gamma \) and \( \delta \),

\[
\text{Cov}(\gamma' Y_t \alpha, \delta' Y_t \beta | F_{t-1}) = v(\alpha' \Sigma_t \beta' \gamma' \Sigma_t \delta + \alpha' \Sigma_t \delta' \gamma' \Lambda_t \delta + \alpha' \Lambda_t \beta' \gamma' \Sigma_t \delta + \alpha' \Lambda_t \delta' \gamma' \Sigma_t \beta + \alpha' \Sigma_t \delta' \gamma' \Lambda_t \beta).
\] \tag{9}

By equation (9), any covariance of interest can be computed. For instance, the portfolio variance with weighing \( \alpha \) is given as \( \alpha' Y_t \alpha \). So the second order dynamic properties of portfolio variances and covariances can be examined directly by equation (9). Another example is the conditional covariance between any two entries of \( Y_t \).
Corollary 1. The conditional variances and covariances for the GCAW\((p, q, r)\) model defined by equations (1)–(4) are given by

\[
\text{Cov}(Y_{ij,t}, Y_{kl,t} \mid \mathcal{F}_{t-1}) = \Lambda_{ik,t} \Sigma_{jl,t} + \Lambda_{il,t} \Sigma_{jk,t} + \Lambda_{jl,t} \Sigma_{ik,t} \\
+ \nu (\Sigma_{il,t} \Sigma_{jk,t} + \Sigma_{jl,t} \Sigma_{ik,t}),
\]

\[
\text{Var}(Y_{ii,t} \mid \mathcal{F}_{t-1}) = 4 \Lambda_{ii,t} \Sigma_{ii,t} + 2 \nu \Sigma_{ii,t}^2,
\]

\[
\text{Var}(Y_{ij,t} \mid \mathcal{F}_{t-1}) = \Lambda_{ii,t} \Sigma_{jj,t} + 2 \Lambda_{ij,t} \Sigma_{ij,t} + \Lambda_{jj,t} \Sigma_{ii,t} + \nu (\Sigma_{ij,t}^2 + \Sigma_{ii,t} \Sigma_{jj,t}),
\]

\[
\text{Cov}(Y_{ii,t}, Y_{kk,t} \mid \mathcal{F}_{t-1}) = 4 \Lambda_{ik,t} \Sigma_{ik,t} + 2 \nu \Sigma_{ik,t}^2,
\]

\[
\text{Cov}(Y_{ii,t}, Y_{kl,t} \mid \mathcal{F}_{t-1}) = 2 \Lambda_{ik,t} \Sigma_{il,t} + 2 \Lambda_{il,t} \Sigma_{ik,t} + 2 \nu \Sigma_{ik,t} \Sigma_{il,t}.
\] (10)

Note that the GCAW\((p, q, r)\) model allows the conditional covariance between two volatilities, and thus their conditional correlation, to be of either the positive or negative sign because the first term in equation (10) can be negative. However, for the CAW\((p, q)\) model, the conditional covariance between two volatilities must be nonnegative since \(\Lambda_{ik,t} = 0\).

For higher order conditional moments, the expressions can be derived by the results in Letac and Massam (2008).

### 2.3 Unconditional moments and stationarity

For the discussion of unconditional moments and stationarity, we derive the VARMA representation for the GCAW\((p, q, r)\) model from the recursions (3) and (4).

Notice that \(y_t\) can be written as

\[
y_t = E(y_t \mid \mathcal{F}_{t-1}) + \nu_t = \lambda_t + \nu \sigma_t + \nu_t,
\] (11)

with \(E(\nu_t) = 0\) and \(E(\nu_t \nu_{t'}) = 0 (s \neq t)\), where \(\nu_t\) is a martingale difference. By substituting equation (5) and \(\nu \sigma_{t-i} = y_{t-i} - \lambda_{t-i} - \nu_{t-i} (i = 0, 1, \ldots, p)\) into equation (7), we have

\[
y_t = \nu c + \sum_{j=1}^{\max(p,q,r)} (\nu \mathcal{A}_j + \mathcal{B}_j + \mathcal{M}_j) y_{t-j} - \sum_{i=1}^{p} \mathcal{B}_i \nu_{t-i} + \nu_t - \sum_{i=1}^{p} \sum_{k=1}^{r} \mathcal{B}_i \mathcal{M}_k y_{t-i-k},
\]
with $\mathcal{A}_j = 0$ for $j > q$, $\mathcal{B}_j = 0$ for $j > p$ and $\mathcal{M}_j = 0$ for $j > r$. Since

$$
\sum_{i=1}^{p} \sum_{k=1}^{r} \mathcal{B}_i \mathcal{M}_k y_{t-i-k} = \sum_{j=2}^{p+r} \sum_{k=1}^{j-1} \mathcal{B}_{j-k} \mathcal{M}_k y_{t-j},
$$

the GCAW($p, q, r$) model can be expressed as a VARMA($\max(p + r, q), p$) model:

$$
y_t = v_c + \sum_{j=1}^{\max(p + r, q)} \left( v \mathcal{A}_j + \mathcal{B}_j + \mathcal{M}_j - \sum_{k=1}^{j-1} \mathcal{B}_{j-k} \mathcal{M}_k \right) y_{t-j} - \sum_{i=1}^{p} \mathcal{B}_i \nu_{t-i} + \nu_t. \tag{12}
$$

The unconditional expectation of $y_t$ follows immediately from the VARMA representation (12), and thus we obtain the conditions for the existence of the unconditional expectation for the GCAW($p, q, r$) model.

**Proposition 3.** The unconditional expectation for the GCAW($p, q, r$) model defined by equations (1)–(4) is given by

$$
E(y_t) = \bar{y} = \left( I_N - \Psi_1 \right)^{-1} v_c, \tag{13}
$$

which is finite if and only if all eigenvalues of the matrix

$$
\Psi_1 = \sum_{j=1}^{\max(p, q, r)} \left[ v \mathcal{A}_j + \mathcal{B}_j + \left( I_N - \sum_{i=1}^{p} \mathcal{B}_i \right) \mathcal{M}_j \right] \tag{14}
$$

are less than 1 in modulus.

For the second moments of $y_t$, we consider the model’s VMA($\infty$) representation. From the VARMA representation (12), the GCAW($p, q, r$) model can be further written as an VMA($\infty$) model (see Lütkepohl, 2005, p. 424):

$$
y_t = \bar{y} + \sum_{i=0}^{\infty} \Phi_i \nu_{t-i}, \tag{15}
$$

with $\Phi_0 = I_N$, and for $i = 1, 2, \ldots$,

$$
\Phi_i = -\mathcal{B}_i + \sum_{j=1}^{i} \left( v \mathcal{A}_j + \mathcal{B}_j + \mathcal{M}_j - \sum_{k=1}^{j-1} \mathcal{B}_{j-k} \mathcal{M}_k \right) \Phi_{i-j}. \tag{16}
$$
Then, provided that the second moments of $y_t$ exist, for $\tau = 0, 1, 2, \ldots$, the lag-$\tau$ autocovariance matrix of $y_t$ has the form

$$\Gamma(\tau) = E [(y_t - \bar{y})(y_{t-\tau} - \bar{y})^\prime] = \sum_{i=0}^{\infty} \Phi_{i+1} E(\nu_i \nu_i^\prime) \Phi_i^\prime,$$

(17)

and when $\tau = 0$, we have the unconditional covariance matrix of $y_t$:

$$\Gamma(0) = \text{Var}(y_t) = E(\nu_t \nu_t^\prime) - \bar{y} \bar{y}^\prime = \sum_{i=0}^{\infty} \Phi_i E(\nu_i \nu_i^\prime) \Phi_i^\prime.$$

(18)

To derive the unconditional second moment for the GCAW($p, q, 1$) model, we prove the following lemma.

**Lemma 1.** Under the GCAW($p, q, 1$) model and the assumption that $E(y_t y_t^\prime)$ exists,

$$\text{vec}[E(\nu_t \nu_t^\prime)] = \Omega (\Omega + I_{N^2})^{-1} (I_{N^2} - M_1 \otimes M_1) \text{vec}[E(y_t y_t^\prime)],$$

where

$$\Omega = \frac{2}{\nu} (D_n^+ \otimes D_n^+) (I_n \otimes K_{nn} \otimes I_n) (D_n \otimes D_n)$$

and $K_{nn}$ is the commutation matrix such that, for any $m \times n$ matrix $X$, $\text{vec}(X^\prime) = K_{mn} \text{vec}(X)$ (see Turkington, 2002, p. 30).

Based on Lemma 1, we obtain the explicit form and thus the existence conditions of the unconditional second moment for the GCAW($p, q, 1$) model.

**Proposition 4.** The unconditional second moment for the GCAW($p, q, 1$) model defined by equations (1)–(4) is given by

$$\text{vec}[E(y_t y_t^\prime)] = (I_{N^2} - \Psi_2)^{-1} \text{vec}(\bar{y} \bar{y}^\prime),$$

(19)

which is finite if and only if all eigenvalues of the matrix

$$\Psi_2 = \left[ \sum_{i=0}^{\infty} (\Phi_i \otimes \Phi_i) \Omega \right] (\Omega + I_{N^2})^{-1} (I_{N^2} - M_1 \otimes M_1)$$

(20)

are less than 1 in modulus.
Proposition 4 implies that the process \( \{ Y_t \} \) under the GCAW\((p, q, 1)\) model is weakly stationary if and only if all eigenvalues of the matrix \( \Psi_2 \) are less than 1 in modulus. In that case, by equations (13), (18) and (19), we obtain the unconditional covariance matrix of \( y_t \), which is given by

\[
\text{vec}[\text{Var}(y_t)] = \text{vec}[E(y_t y_t')] - \text{vec}(\bar{y} \bar{y}') = \left( \Psi_2^{-1} - I_{N^2} \right)^{-1} \text{vec}(\bar{y} \bar{y}').
\]

Moreover, by combining the results of Lemma 1 and Proposition 4, we obtain the unconditional covariance matrix of the martingale difference \( \nu_t \), which is given by

\[
\begin{align*}
\text{vec}[E(\nu_t \nu_t')] &= \Omega \left( \Omega + I_{N^2} \right)^{-1} \left( I_{N^2} - M_1 \otimes M_1 \right) \left( I_{N^2} - \Psi_2 \right)^{-1} \text{vec}(\bar{y} \bar{y}') \\
&= \Omega \left[ \left( I_{N^2} - M_1 \otimes M_1 \right)^{-1} \left( \Omega + I_{N^2} \right) - \sum_{i=0}^{\infty} (\Phi_i \otimes \Phi_i) \Omega \right] \left( \Omega + I_{N^2} \right)^{-1} \text{vec}(\bar{y} \bar{y}').
\end{align*}
\]

and can be used to compute the lag-\( \tau \) autocovariance matrix of \( y_t \) as given by equation (17).

A special case is the CAW\((p, q)\) model, where \( M_1 = 0 \). Since \( \Phi_0 = I_N \), in this case, equation (19) can be further reduced to

\[
\begin{align*}
\text{vec}[E(y_t y_t')] &= \left\{ I_{N^2} - \left[ \sum_{i=0}^{\infty} (\Phi_i \otimes \Phi_i) \Omega \right] \right\} \left( \Omega + I_{N^2} \right)^{-1} \text{vec}(\bar{y} \bar{y}') \\
&= \left\{ I_{N^2} - \left[ \Omega + \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) \Omega \right] \right\} \left( \Omega + I_{N^2} \right)^{-1} \text{vec}(\bar{y} \bar{y}') \\
&= (\Omega + I_{N^2}) \left[ \Omega + I_{N^2} - \Omega - \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) \Omega \right] \left( \Omega + I_{N^2} \right)^{-1} \text{vec}(\bar{y} \bar{y}') \\
&= (\Omega + I_{N^2}) I_{N^2} - \sum_{i=1}^{\infty} (\Phi_i \otimes \Phi_i) \Omega \left( \Omega + I_{N^2} \right)^{-1} \text{vec}(\bar{y} \bar{y}'),
\end{align*}
\]

which is consistent with the result in Golosnoy et al. (2012).

Another special case is the WAR(1) model, where \( \mathcal{A}_1 = \cdots = \mathcal{A}_q = \mathcal{B}_1 = \cdots = \mathcal{B}_p = 0 \). In this case, we derive the following results.

**Corollary 2.** The unconditional first and second moments for the WAR(1) model (Gouriéroux et al.,
are given by
\[ E(y_t) = \bar{y} = (I_N - M_1)^{-1} \nu, \]
\[ \text{vec}[E(y_t y_t')] = (I_{N^2} - M_1 \otimes M_1)^{-1} (\Omega + I_{N^2}) (I_{N^2} - M_1 \otimes M_1) \text{vec}(\bar{y} \bar{y}'), \]
which are finite if and only if all eigenvalues of the matrix \( M_1 = L_n(M_1 \otimes M_1) D_n \) are less than 1 in modulus.

By equation (21), the unconditional covariance matrix of \( y_t \) is given by
\[ \text{vec}[\text{Var}(y_t)] = (I_{N^2} - M_1 \otimes M_1)^{-1} \Omega (I_{N^2} - M_1 \otimes M_1) \text{vec}(\bar{y} \bar{y}'). \]

### 2.4 Parameter estimation

The parameters of the GCAW(\( p, q, r \)) model, \( \theta = (\nu, \text{vech}(C)', \text{vec}(B_1)', \ldots, \text{vec}(B_p)', \text{vec}(A_1)', \ldots, \text{vec}(A_q)', \text{vec}(M_1)', \ldots, \text{vec}(M_r)') \), can be estimated by the maximum likelihood estimation. By equation (2), the log-likelihood function is given as
\[
\ell_T(\theta) = - T \times \left[ \frac{v n}{2} \log 2 + \frac{n(n - 1)}{4} \log \pi + \sum_{i=1}^{n} \log \Gamma \left( \frac{v + 1 - i}{2} \right) \right] \\
+ \sum_{i=1}^{T} \left\{ - \frac{v}{2} \log \det \Sigma_i + \frac{v - n - 1}{2} \log \det Y_i - \frac{1}{2} \text{tr} \left[ \Sigma_i^{-1} (Y_i + \Lambda_i) \right] \\
+ \log {}_0F_1 \left( \frac{v}{2}; \frac{1}{4} \Sigma_i^{-1} \Lambda_i \Sigma_i^{-1} Y_i \right) \right\}, \tag{22}
\]
where \( {}_0F_1 \) is the hypergeometric function of matrix argument, which has no closed form expression, but can be efficiently approximated by the algorithm developed by Koev and Edelman (2006). The ML-estimates are obtained by maximizing the log-likelihood function numerically. We follow Golosnoy et al. (2012) to use a 'bottom-up' model estimating strategy in order to provide good starting values for the optimization algorithm. That is, we estimate the model in a sequence of order \((0, 1, 0), (1, 1, 0), (1, 2, 0), \ldots, \) where the estimates for the previous order are used as starting values. Then, for each pair \((p, q)\), we start with the order \((p, q, 0)\) and increase successively the order \(r\). For each order \((p, q, r)\), we use different starting values to check if a local optimum is achieved. To restrict \( C_{11}, \ldots, C_{mn}, A_{11,1}, \ldots, A_{11,q}, B_{11,1}, \ldots, B_{11,p}, M_{11,1}, \ldots, M_{11,r} > 0 \)
and $v > n - 1$, $\sqrt{C_{ii}}$, $\sqrt{A_{11,ij}}$, $\sqrt{B_{11,ij}}$, $\sqrt{M_{11,ik}}$ and $\sqrt{v-n+1}$ are estimated instead.

Although the estimation is computationally intensive, it is still feasible. For example, the ML-estimation of a GCAW(2, 2, 1) model for the RCOV matrix of five assets with 141 parameters takes of the order of 10 minutes on an Intel Core i5 3.3GHz processor using MATLAB on Windows 7.

2.5 Model identification and diagnostic checking

For model identification, we use Schwarz’s (1978) Bayesian information criterion (BIC) to choose the optimal order $(p, q, r)$. The BIC of the GCAW$(p, q, r)$ model is given as

$$\text{BIC} = -2\ell_T(\hat{\theta}) + \text{dim}(\theta) \times \log T,$$  

(23)

where $\hat{\theta}$ is the ML-estimates of $\theta$, $\ell_T(\cdot)$ is the log-likelihood function given in equation (22) and $\text{dim}(\theta) = n(n+1)/2 + (p+q+r)n^2 + 1$ is the number of parameters.

For model diagnostic checking, we define the standardized residual vector as

$$e^*_t = \text{Var}(y_t | F_{t-1})^{-1/2} [y_t - E(y_t | F_{t-1})],$$  

(24)

where $E(y_t | F_{t-1})$ and $\text{Var}(y_t | F_{t-1})$ are given in equations (7) and (8), respectively, with $\theta = \hat{\theta}$, and $\text{Var}(y_t | F_{t-1})^{-1/2}$ is the inverse of the Cholesky factor of matrix $\text{Var}(y_t | F_{t-1})$. If the model adequately describes the temporal dependence in elements of $y_t$, the standardized residuals $e^*_i$, the elements of the vector $e^*_t$, should be serially uncorrelated. Therefore, a portmanteau test statistic (Hosking, 1980; Li and McLeod, 1981) constructed from the standardized residuals $e^*_i$ can be used to check the adequacy of the fitted models.
2.6 Forecasting

Given the parameter estimates, we obtain the $h$-step-ahead forecasts of $y_t$, $\hat{y}_t(h)$ ($h = 1, 2, \ldots$), by recursion. By equation (7), it can be shown that

$$
\hat{y}_t(h) = E(y_{t+h} | F_t) = E [E(y_{t+h} | F_{t+h-1}) | F_t]
$$

$$
= vc + \sum_{i=1}^{p} vB_i E(\sigma_{t+h-i} | F_t) + \sum_{j=1}^{\max(q,r)} (vA_j + M_j) E(y_{t+h-j} | F_t), \quad (25)
$$

where, for $\tau = 1, 2, \ldots$,

$$
E(y_{t+h-\tau} | F_t) = \begin{cases} 
    y_{t+h-\tau}, & \text{if } \tau \geq h \\
    \hat{y}_t(h - \tau), & \text{if } \tau < h,
\end{cases}
$$

$$
E(\sigma_{t+h-\tau} | F_t) = \begin{cases} 
    \sigma_{t+h-\tau}, & \text{if } \tau \geq h - 1 \\
    c + \sum_{i=1}^{p} B_i E(\sigma_{t+h-i} | F_t) + \sum_{j=1}^{q} A_j E(y_{t+h-j} | F_t), & \text{if } \tau < h - 1.
\end{cases}
$$

By applying equation (25) recursively, we compute $\hat{y}_t(1)$, $\hat{y}_t(2), \ldots$. Then, the $h$-step-ahead forecasts of the RCOV matrix $Y_t$, $\hat{Y}_t(h) = E(Y_{t+h} | F_t)$ ($h = 1, 2, \ldots$), can be obtained by substituting the elements of vector $\hat{y}_t(h)$ into a symmetric matrix.

3 Simulation studies

For simulating the GCAW model, we need to generate samples from the noncentral Wishart distribution. This relies on the following factorization (see Muirhead, 1982, p. 441). For an $n \times n$ random matrix $Y \sim W_n(v, \Lambda, \Sigma)$ with integer degrees of freedom $v \geq n$, $n \times n$ symmetric and positive definite noncentrality matrix $\Lambda$ and scale matrix $\Sigma$, since it is symmetric and positive definite, it can always be factorized as $Y = X'X$ such that $X$ is a $v \times n$ random matrix, following a matrix-variate normal distribution (Gupta and Nagar, 1999) $X \sim N_{v,n}(N, \Sigma \otimes I_v)$, where $N$ is a $v \times n$ matrix such that $\Lambda = N'N$, and $I_v$ is the identity matrix with dimension $v$. By the properties of the matrix-variate normal distribution, matrix $X$ can be further written as $X = N + Z\Sigma^{1/2}$, where $Z \sim N_{v,n}(0, I_v \otimes I_v)$ is a matrix containing uncorrelated standard normal random numbers.
The simulation procedure is as follow:

1. Given the parameters and the lagged values of $Y_t$ and $\Sigma_t$, calculate $\Lambda_t$ and $\Sigma_t$ by equations (3) and (4), with $Y_0 = Y_{-1} = Y_{-2} = \cdots = \Sigma_0 = \Sigma_{-1} = \Sigma_{-2} = \cdots = I_n$.

2. Perform the Cholesky decompositions of $\Lambda_t$ and $\Sigma_t$ such that

$$\Lambda_t = \bar{N}_t \bar{N}_t'$$
and
$$\Sigma_t = D_t' D_t.$$

3. Construct the $v \times n$ mean matrix as

$$N_t = \begin{bmatrix} \bar{N}_t \\ O_{v-n,n} \end{bmatrix},$$

where $O_{v-n,n}$ is the $(v-n) \times n$ zero matrix.

4. Generate a $v \times n$ random matrix $Z$ such that its elements are uncorrelated standard normal random numbers.

5. Obtain the simulated RCOV matrix as

$$Y_t = (N_t + ZD_t)' (N_t + ZD_t).$$

Two simulation experiments are carried out. In each experiment, 100 samples of size $T = 600$ are simulated, and the first 100 observations are deleted to remove the effect of letting $Y_0 = Y_{-1} = Y_{-2} = \cdots = \Sigma_0 = \Sigma_{-1} = \Sigma_{-2} = \cdots = I_n$. This leaves 100 sequences of 500 RCOV matrices for three stocks. In the first experiment, the parameters take the values of model 1 (i.e., a GCAW(1, 2, 1) model) shown in the upper panel of Table 1. For each sample, we fit the GCAW model with orders $(p, q, r)$ ranging from $(0, 0, 0)$ to $(2, 2, 2)$. The BIC given in equation (23) correctly chooses the order $(1, 2, 1)$ in all samples. Moreover, the likelihood ratio test that involves the CAW(1, 2) as the restricted model and the GCAW(1, 2, 1) as the unrestricted model (i.e., $H_0 : M_1 = 0$) correctly rejects the CAW(1, 2) model in all samples, all with $p$-value close to 0. These results illustrate that the GCAW model with time-varying noncentrality matrix outperforms the otherwise identical model with zero noncentrality matrix when the true process involves noncentrality of the Wishart distribution.
In the second experiment, the parameters take the values of model 2 (i.e., a CAW(1, 2) model) shown in the lower panel of Table 1. Again, for each sample, we fit the GCAW model with orders \((p, q, r)\) ranging from \((0, 0, 0)\) to \((2, 2, 2)\). The BIC correctly chooses the order \((1, 2, 0)\) in all samples. We also perform the above likelihood ratio test. There are 85 out of 100 samples in which the test indicates that the noncentrality parameter \(M_1\) in the GCAW(1, 2, 1) model is insignificant at the 5\% significance level, and thus reduces to the true CAW(1, 2) model.

4 Empirical applications

4.1 Data

Consider the data set of daily RCOV matrices for five stocks traded at the New York Stock Exchange: American Express (AXP), Citigroup (C), General Electric (GE), Home Depot (HD) and International Business Machines (IBM), starting at 14 December 2000 and ending on 31 December 2009, with total 2274 observations. The data set is also evaluated by Golosnoy et al. (2012) with the same sampling period but a different RCOV measure. In Golosnoy et al. (2012), the daily RCOV matrices are obtained by averaging over 30 subsampling subgrids per day. We refine this RCOV measure by using the averaged realized volatility matrix (ARVM) estimator (without thresholding) proposed by Wang and Zou (2010). The market was opened between 9:30 a.m. and 4:00 p.m. during the sampling period, and the first 30 minutes are deleted to remove the opening effects. For each trading day, sets of five-minute grid are selected as pre-sampling frequencies. Given the tick-by-tick price data, an RCOV matrix is constructed using the previous tick method according to each pre-sampling frequency. Then, we take the average of the constructed RCOV matrices in order to reduce market microstructure noise and exploit the data richness more efficiently. Since the ARVM estimator does not guarantee positive semi-definiteness, we also check all the matrices which confirm that all such estimates are positive definite. This gives a sequence of 2274 matrices of \(Y_t (t = 1, \ldots, 2274)\) of size 5 by 5.

Figure 1 plots the time series of all variances and covariances. These plots reveal that there are two sub-periods during the sampling period in which both the value and volatility of the elements of \(Y_t\) increase significantly: the aftermath of the dot-com bubble during the early 2000s and the U.S. subprime mortgage crisis starting in 2008. Summary statistics of the RCOVs \(Y_{ij}\)'s provided in
Table 2 illustrate that the variances and covariances are highly positively skewed and possess heavy
tails. To investigate the long-run dependence in the time series, following Andersen et al. (2003),
we also calculate the Geweke and Porter-Hudak (1983) log-periodogram regression estimates of
the fractional integration parameter \( d \) for each variance and covariance based on the \( m = T^{4/5} = 485 \)
lowest-frequency periodogram ordinates, which are shown in the last column of Table 2. All
estimates of \( d \) are significantly greater than zero when judged by the asymptotic standard error of
\( \pi \times (24m)^{-1/2} = 0.0291 \), and most of them are less than one half. This indicates that the time series
of the RCOVs are strongly serially correlated but still stationary. All these results are consistent
with those in Golosnoy et al. (2012).

Notice that there are a handful of RCOV estimators which ensure positive semi-definiteness,
including Barndorff-Nielsen et al. (2011), Christensen et al. (2010), and Shephard and Xiu (2014).
For cross-validation and justifying the use of the ARVM estimator, we also obtain the daily RCOV
matrices by using the multivariate realized kernel estimator proposed by Barndorff-Nielsen et al.
(2011). Summary statistics reported in Table 3 show that the RCOVs given by the two different
RCOV measures share similar characteristics. Moreover, previous studies (Zhang et al., 2005;
Wang and Zou, 2010; Tao et al., 2011) showed that the ARVM estimator is consistent under some
regularity conditions. For the sake of comparability with the results in Golosnoy et al. (2012), we
use the ARVM estimator (without thresholding) as the RCOV measure.

4.2 Estimation results

We now fit the GCAW model to the full sample data of daily RCOV matrices, with orders \((p, q, r)\)
ranging from \((0, 0, 0)\) to \((3, 3, 2)\). In particular, the GCAW\((0, 0, 0)\) model represents the central
Wishart distribution \(W_n(v, CC')\); the GCAW\((0, 0, r)\) model represents the WAR\((r)\) model pro-
posed by Gouriéroux et al. (2009); and the GCAW\((p, q, 0)\) model represents the CAW\((p, q, 0)\) model
proposed by Golosnoy et al. (2012). Table 4 reports the maximum values of the log-likelihood
function given in equation (22), BIC given in equation (23), the maximum eigenvalues of the ma-
trices \(W_1\) and \(W_2\) given in equations (14) and (20), and the results of the portmanteau test using
100 lags of the standardized residuals as described in Section 2.5.

The best specification based on the BIC is the GCAW\((2, 2, 1)\), in which its ML-estimates are
given in the lower panel of Table 5. Figure 2 plots the fitted values of all variances and covariances
under the GCAW(2, 2, 1) model. The period covered is from 1 July 2008 to 30 June 2009, during which the market is influenced by the U.S. subprime mortgage crisis with a very high volatility. It can be seen that this high-volatility period can also be well captured by the model. The maximum eigenvalues of the matrices $\Psi_1$ and $\Psi_2$ imply that RCOVs under the GCAW(2, 2, 1) model have finite unconditional first and second moments, and thus are weakly stationary. The diagnostic check result reveals that the GCAW(2, 2, 1) model adequately accounts for the temporal dependence in the RCOVs, with all residual series pass the portmanteau test at the 1% significance level. Although RCOVs under the WAR model are weakly stationary, the WAR model is the worst according to both the BIC and diagnostic check result among the WAR, CAW and GCAW models. All residual series of all WAR models fail the portmanteau test. This suggests that the constant scale matrix fails to explain the temporal dependence in the RCOVs, leading to a poor fit. On the other hand, for each pair $(p, q)$, despite the fact that the GCAW($p$, $q$, 1) model has more parameters than the CAW($p$, $q$) model, the BIC favors the GCAW($p$, $q$, 1) specification over its CAW counterpart. It is because the inclusion of the time-varying noncentrality matrix significantly increases the maximum value of the log-likelihood function, indicating a much better fit. For all the CAW specifications, at least one of the maximum eigenvalues of the matrices $\Psi_1$ and $\Psi_2$ obtained are greater than one, whereas those obtained for the GCAW($p$, $q$, 1) specifications, with $p$ and $q$ greater than one, are less than one. It implies that RCOVs under all of the GCAW($p$, $q$, 1) specifications are weakly stationary, while those under all CAW specifications are not. This suggests that the inclusion of the time-varying noncentrality matrix substantially reduces the persistence in the scale matrix of the Wishart distribution and has significant effects on the dynamic structure of the model, making the model weakly stationary. Moreover, the GCAW model with time-varying noncentrality matrix generally provides a better diagnostic check result than the CAW model. The result of the portmanteau test suggests that all CAW models cannot fully capture the temporal dependence in the RCOVs, which is consistent with the finding in Golosnoy et al. (2012). The GCAW model’s domination in terms of accounting for serial correlation in the RCOVs may be explained by the fact that the GCAW model gives more flexible conditional covariances of the RCOVs than the CAW model, and thus describes the dependence in the RCOVs in a better way.

To investigate if the time-varying noncentrality matrix and scale matrix are necessary when using the Wishart distribution to model the RCOV matrix, for each GCAW($p$, $q$, $r$) model with $r = 1, 2$, two sets of the likelihood ratio test are performed. The first set of the likelihood ratio test involves the CAW($p$, $q$) model as the restricted model, with the null hypothesis of a zero
noncentrality matrix (all $M_k$’s = 0, $k = 1, \ldots, r$). Rejection of the null hypothesis indicates that the time-varying noncentrality matrix is important (i.e., the autoregressive property of high-frequency returns is significant). The second set of the likelihood ratio test involves the WAR($r$) model as the restricted model, with the null hypothesis of a constant scale matrix (all $A_i$’s and $B_j$’s = 0, $i = 1, \ldots, p; j = 1, \ldots, q$). Rejection of the null hypothesis indicates that the time-varying scale matrix is important (i.e., the conditional heteroscedasticity of returns is significant). The likelihood ratio test statistic under the null hypothesis is chi-squared distributed with degrees of freedom equals to the difference between the number of parameter in the restricted and unrestricted model. Table 6 reports the test statistic, critical value at the 5% significance level and $p$-value of the likelihood ratio tests. All $p$-values are close to 0, revealing that both the autoregressive property of high-frequency returns and the conditional heteroscedasticity of returns are significant. It gives evidence that, when using the Wishart distribution to model the RCOV matrix, both the noncentrality matrix and scale matrix should be time-varying. Moreover, for each order ($p, q, r$), the second likelihood ratio statistic is more significant than the first one. This suggests that while both sources of dynamics are important, the conditional heteroscedasticity of returns contributes more to the dynamics of the RCOVs.

4.3 Forecasting results

We now compare the out-of-sample-forecast performance among various GCAW specifications. Again, GCAW models with orders ($p, q, r$) ranging from (0, 0, 0) to (3, 3, 2) are fitted. Each model is re-estimated daily with a moving window sample consisting of the past 500 trading days. For each day in the out-of-sample period, based on the updated parameter estimates, the $h$-step-ahead forecasts of the RCOV matrix $\hat{Y}_t(h)$ as described in Section 2.6 are computed, with forecast horizons $h = \{1, 5, 10\}$ days. Then, the forecasts are compared with the ex-post realization of the RCOV matrix $Y_{t+h}$.

For our forecast experiment, we consider two out-of-sample windows which cover the normal and crisis periods in order to obtain a balanced assessment of the forecast performance of the GCAW model. Following Golosnay et al. (2012), we select the period from 2 July 2007 to 30 June 2008 (252 trading days) to be the first window, where the volatility is comparably low, and thus it represents the normal period. The second window covers the period from 1 July 2008 to 30 June 2009 (252 trading days), during which the market is influenced by the U.S. subprime mortgage
crisis with a very high volatility, and thus it represents the crisis period.

To evaluate the forecasting accuracy for a given model, we follow Ledoit et al. (2003) and Golosnoy et al. (2012) to measure the forecast error by the root-mean-square error based on the Frobenius norm. Then, the out-of-sample-forecast performance among various GCAW specifications are compared by the average forecast error for 252 forecasts, given by

\[ FN_h = \frac{1}{T_h} \sum_t \left\| Y_{t+h} - \hat{Y}_t(h) \right\|_F = \frac{1}{T_h} \sum_t \left\{ \sum_{i,j} \left[ Y_{ij,t+h} - \hat{Y}_{ij,t}(h) \right]^2 \right\}^{1/2}, \]

where \( T_h \) is the number of forecast periods and equals to 252. For other measures to access the forecasting accuracy of multivariate volatility models, please refer to Laurent et al. (2009).

Table 7 reports the results of forecasting accuracy for different GCAW models. For the normal period, the GCAW(3, 3, 2) model yields the most accurate forecasts at the 1-day horizon, whereas the GCAW(2, 3, 2) model is the best one for 5-day and 10-day forecasts. Again, among the WAR, CAW and GCAW models, the WAR models have the worst forecasting performance, which only perform slightly better than the time-invariant central Wishart distribution at the 5-day and 10-day forecast horizons. Generally, the GCAW models with time-varying noncentrality matrix outperform their CAW counterpart at various forecast horizons. This indicates the importance of including both time-varying noncentrality matrix and scale matrix of the Wishart distribution in modeling the RCOV matrix.

For the crisis period, as we may expect, the forecasting accuracy of all models substantially deteriorates, with a large increase in the average forecast error across all models and time horizons. Among all models, the GCAW(2, 3, 1) model is the best one for 1-day forecasts. Interestingly, the WAR(3) (i.e., GCAW(0, 0, 3)) model gives the most accurate forecasts at the 5-day and 10-day horizons. It may be because the GCAW models with time-varying scale matrix tend to overstate the persistence of volatility clustering after a significant shock, whereas the WAR models give more stable forecasts at a long time horizon. Nevertheless, the forecasting accuracy of the GCAW models with time-varying scale matrix is improved when we increase the lag orders, especially the lag order of noncentrality matrix. Again, the GCAW models with time-varying noncentrality matrix outperform their CAW counterpart at various forecast horizons.
5 Conclusions and discussions

In this paper, we propose the generalized conditional autoregressive Wishart (GCAW) model for the analysis of dynamics of realized covariance (RCOV) matrices of asset returns. The model represents a generalization of both the Wishart autoregressive (WAR) model (Gouriéroux et al., 2009) and the conditional autoregressive Wishart (CAW) model (Golosnoy et al., 2012), and pools together their strengths by incorporating two different sources of dynamics of RCOVs, namely, autoregressive high-frequency returns and conditional heteroscedastic returns. As such, it accounts for complex temporal and cross-sectional dependence in multivariate RCOVs. There are several advantages in using the Wishart distribution to model the RCOV matrix. Firstly, the maximum likelihood estimation can be implemented easily by virtue of the closed form density function. Moreover, analytical conditional moments are available, which are useful especially for model diagnostics. Stationarity conditions can also be derived in a straightforward manner. Most importantly, symmetry and positive definiteness of RCOV matrices are guaranteed without imposing any parametric restriction.

Both the simulation studies and empirical applications demonstrate that it is necessary to include both time-varying noncentrality matrix and scale matrix of the Wishart distribution for the analysis of RCOV matrices. The empirical example of daily RCOV matrices from the New York Stock Exchange shows that the GCAW models with time-varying noncentrality matrix and scale matrix outperform the WAR and CAW models in terms of goodness of fit and forecasting. Furthermore, the standardized residuals show no significant serial correlation indicating that the proposed model can adequately capture the highly complex dynamic nature in the RCOV matrix, whereas the CAW only partially removes the observed temporal dependence in the RCOVs, which is also pointed out in Golosnoy et al. (2012).

From the computational point of view, the model estimation can be demanding when the number of assets $n$ increases, since the GCAW($p, q, r$) model involves $n(n + 1)/2 + (p + q + r)n^2 + 1$ parameters. However, note that the GCAW model is directly fitted to $n(n + 1)/2$ RCOVs, whereas the corresponding multivariate GARCH models are estimated based on $n$ returns only. The number of observations per parameter is much larger for the GCAW model than that for the similarly parameterized GARCH models. Therefore, the curse of dimensionality problem appears to be less acute for the GCAW model. Those who are interested in conditional variance models that challenges the curse of dimensionality can refer to Barigozzi et al. (2014) on realized volatilities using...
a multiplicative error model (MEM) specification with a locally stationary factor structure. To consider further reduce the number of parameters, particularly in high dimensional GCAW models, one can consider to incorporate a factor-type GARCH structure similar to the models proposed by Sentana et al. (2008) who developed a factor-GARCH models estimated based on returns. Application of this kind can be found in risk management and portfolio allocation. Such interesting problem and its applications will be considered in the future.

Acknowledgments

The authors thank two anonymous referees and the associate editor for their helpful and very constructive comments.

Appendix. Proofs

Proof for Proposition 1. Under the Wishart distribution in equation (1), by Theorem 4.4 in Magnus and Neudecker (1979), we have

\[ E(Y_t | F_{t-1}) = \Lambda_t + v\Sigma_t, \]  
\[ \text{Var}[\text{vec}(Y_t) | F_{t-1}] = (I_{n^2} + K_{nn})[v(\Sigma_t \otimes \Sigma_t) + \Sigma_t \otimes \Lambda_t + \Lambda_t \otimes \Sigma_t], \]

where \( K_{nn} \) is the commutation matrix such that, for any \( m \times n \) matrix \( X \), \( \text{vec}(X^\prime) = K_{nn} \text{vec}(X) \). By equations (5), (6) and (27), we can see that

\[ E(y_t | F_{t-1}) = E[\text{vech}(Y_t) | F_{t-1}] \]
\[ = \text{vech}[E(Y_t | F_{t-1})] \]
\[ = \text{vech}(\Lambda_t + v\Sigma_t) \]
\[ = \lambda_t + v\sigma_t \]
\[ = \text{vech} \left( \sum_{i=1}^{p} vB_i \sigma_{t-i} + \sum_{j=1}^{\max(q,r)} (vA_j + M_j)y_{t-j} \right). \]
Define $\mathcal{N}_n = (I_n \otimes \mathcal{K}_m)/2$ (see Turkington, 2002, p. 42-44). Since $\mathcal{N}_n(A \otimes A) = \mathcal{N}_n(A \otimes A)\mathcal{N}_n'$ and $\mathcal{N}_n(A \otimes B + B \otimes A) = \mathcal{N}_n(A \otimes B + B \otimes A)\mathcal{N}_n'$ for any $n \times n$ matrices $A$ and $B$, $\mathcal{L}_n\mathcal{N}_n = \mathcal{D}_n^+$ and $y_t = \text{vech}(Y_t) = \mathcal{L}_n\text{vec}(Y_t)$, by equation (28), we obtain

$$\text{Var}(y_t | \mathcal{F}_{t-1}) = \text{Var}[\mathcal{L}_n\text{vec}(Y_t) | \mathcal{F}_{t-1}]$$

$$= \mathcal{L}_n(I_n \otimes \mathcal{K}_m)[v(\Sigma_i \otimes \Sigma_i) + \Sigma_i \otimes \Lambda_i + \Lambda_i \otimes \Sigma_i]\mathcal{L}_n'$$

$$= 2\mathcal{L}_n\mathcal{N}_n[v(\Sigma_i \otimes \Sigma_i) + \Sigma_i \otimes \Lambda_i + \Lambda_i \otimes \Sigma_i]\mathcal{L}_n'$$

$$= 2\mathcal{L}_n\mathcal{N}_n[v(\Sigma_i \otimes \Sigma_i) + \Sigma_i \otimes \Lambda_i + \Lambda_i \otimes \Sigma_i]\mathcal{N}_n'\mathcal{L}_n'$$

$$= 2\mathcal{D}_n^+[v(\Sigma_i \otimes \Sigma_i) + \Sigma_i \otimes \Lambda_i + \Lambda_i \otimes \Sigma_i](\mathcal{D}_n^+)'$$

which completes the proof.

**Proof for Proposition 2.** Since $(A \otimes B)(C \otimes D) = AC \otimes BD$, and for any $n \times n$ matrix $X$, $n \times 1$ vectors $a$ and $b$, $a'Xb = (b \otimes a')\text{vec}(X)$ and $\mathcal{K}_m(a \otimes b) = b \otimes a$ (see Turkington, 2002, p. 8, 11, 31), by equation (28), we obtain

$$\text{Cov}(\gamma'Y_t\alpha, \delta'Y_t\beta | \mathcal{F}_{t-1})$$

$$= \text{Cov}[(\alpha' \otimes \gamma')\text{vec}(Y_t), (\beta' \otimes \delta')\text{vec}(Y_t) | \mathcal{F}_{t-1}]$$

$$= (\alpha' \otimes \gamma')(I_n \otimes \mathcal{K}_m)[v(\Sigma_i \otimes \Sigma_i) + \Sigma_i \otimes \Lambda_i + \Lambda_i \otimes \Sigma_i](\beta \otimes \delta)$$

$$= (\alpha' \otimes \gamma')(v(\Sigma_i \otimes \Sigma_i) + \Sigma_i \otimes \Lambda_i + \Lambda_i \otimes \Sigma_i)(\beta \otimes \delta)$$

$$+ (\alpha' \otimes \gamma')\mathcal{K}_m[v(\Sigma_i \otimes \Sigma_i) + \Sigma_i \otimes \Lambda_i + \Lambda_i \otimes \Sigma_i](\beta \otimes \delta)$$

$$= v(\alpha'\Sigma_i\beta \otimes \gamma'\Sigma_i\delta) + \alpha'\Sigma_i\beta \otimes \gamma'\Lambda_i\delta + \alpha'\Lambda_i\beta \otimes \gamma'\Sigma_i\delta$$

$$+ (\alpha' \otimes \gamma')\mathcal{K}_m[v(\Sigma_i\beta \otimes \Sigma_i\delta) + \Sigma_i\beta \otimes \Lambda_i\delta + \Lambda_i\beta \otimes \Sigma_i\delta]$$

$$= v(\alpha'\Sigma_i\beta \gamma'\Sigma_i\delta + \alpha'\Sigma_i\beta \gamma'\Lambda_i\delta + \alpha'\Lambda_i\beta \gamma'\Sigma_i\delta$$

$$+ (\alpha' \otimes \gamma')v(\Sigma_i\delta \otimes \Sigma_i\delta) + \Lambda_i\delta \otimes \Sigma_i\beta + \Sigma_i\delta \otimes \Lambda_i\beta$$

$$= v(\alpha'\Sigma_i\beta \gamma'\Sigma_i\delta + \alpha'\Sigma_i\delta \gamma'\Sigma_i\beta) + \alpha'\Sigma_i\delta \gamma'\Lambda_i\delta$$

$$+ \alpha'\Lambda_i\beta \gamma'\Sigma_i\delta + \alpha'\Lambda_i\delta \gamma'\Sigma_i\beta + \alpha'\Sigma_i\delta \gamma'\Lambda_i\beta,$$

which completes the proof.

**Proof for Corollary 1.** The conditional covariance $\text{Cov}(Y_{ij,t}, Y_{kl,t} | \mathcal{F}_{t-1})$ can be obtained by using Proposition 2 with $\alpha = e_j$, $\beta = e_l$, $\gamma = e_i$ and $\delta = e_k$, where $e_i$ is the vector with zero entries.
except the $i$-th entry which equals 1. The other results of this corollary follow immediately.

**Proof for Proposition 3.** By taking expectation on both sides of equation (12), since $E(\nu_i) = 0$, we have

$$E(y_i) = \text{ve}c + \sum_{j=1}^{\max(p+q,r)} \left( vA_j + B_j + M_j - \sum_{k=1}^{j-1} B_{j-k} M_k \right) E(y_i)$$

$$= \text{ve}c + \sum_{j=1}^{\max(p,q,r)} \left( vA_j + B_j + M_j - \sum_{k=1}^{j-1} B_{j-k} M_k \right) E(y_i)$$

$$= \text{ve}c + \sum_{j=1}^{\max(p,q,r)} \left( vA_j + B_j + M_j - \sum_{k=1}^{j} B_k M_j \right) E(y_i)$$

$$= \text{ve}c + \sum_{j=1}^{\max(p,q,r)} \left( vA_j + B_j + M_j - \sum_{k=1}^{p} B_k M_j \right) E(y_i)$$

$$= \text{ve}c + \Psi_1 E(y_i),$$

which can be solved for $E(y_i)$ to obtain equation (13) if and only if all eigenvalues of the matrix $\Psi_1$ are less than 1 in modulus.

**Proof for Lemma 1.** By equation (8) and the identity $\text{ve}c(ABC) = (C' \otimes A)\text{ve}c(B)$ (see Turkington, 2002, p. 11), we have

$$\text{ve}c [\text{Var}(y_i \mid \mathcal{F}_{i-1})] = \frac{2}{v} (D_n^+ \otimes D_n^+) \text{ve}c [v^2 (\Sigma_i \otimes \Sigma_i) + v (\Sigma_i \otimes \Lambda_i) + v (\Lambda_i \otimes \Sigma_i)]. \tag{29}$$

Since, for any $n \times n$ matrices $A$ and $B$, $\text{ve}c(A \otimes B) = (I_n \otimes K_{nn} \otimes I_n)[\text{ve}c(A) \otimes \text{ve}c(B)]$ (see Turkington, 2002, p. 34) and $\text{ve}c(A) \otimes \text{ve}c(B) = (D_n \otimes D_n)\text{ve}c(\text{vech}(B)[\text{vech}(A)]')$, we can write

$$\text{ve}c[v^2 (\Sigma_i \otimes \Sigma_i) + v (\Sigma_i \otimes \Lambda_i) + v (\Lambda_i \otimes \Sigma_i)]$$

$$= (I_n \otimes K_{nn} \otimes I_n)[v^2 [\text{ve}c(\Sigma_i) \otimes \text{ve}c(\Sigma_i)] + v [\text{ve}c(\Sigma_i) \otimes \text{ve}c(\Lambda_i)] + v [\text{ve}c(\Lambda_i) \otimes \text{ve}c(\Sigma_i)]]$$

$$= (I_n \otimes K_{nn} \otimes I_n)(D_n \otimes D_n)\text{ve}c(v^2 \sigma_i \sigma_i' + v \lambda_i \sigma_i' + v \sigma_i \lambda_i). \tag{30}$$

Then, by substituting equation (30) into equation (29) and then taking expectation on both sides, we obtain

$$\text{ve}c[E[\text{Var}(y_i \mid \mathcal{F}_{i-1})]] = \Omega \text{ve}c[E(v^2 \sigma_i \sigma_i' + v \lambda_i \sigma_i' + v \sigma_i \lambda_i)]. \tag{31}$$
Since \( \text{Var}(y_t | F_{t-1}) = E(y_t | F_{t-1}) - E(y_t | F_{t-1})E(y_t | F_{t-1}) \), by the law of iterated expectations and equation (7), we have

\[
E[\text{Var}(y_t | F_{t-1})] = E(y_t y_t') - E[(\lambda_t + v \sigma_t) (\lambda_t + v \sigma_t')].
\] (32)

By the fact that \( E(v_t v_t') = E(y_t y_t') - E[(\lambda_t + v \sigma_t) (\lambda_t + v \sigma_t')] \) (see equation (11)) and equations (31) and (32), we can write

\[
\text{vec}[E(v_t v_t')] = \text{vec}[E[\text{Var}(y_t | F_{t-1})]] = \Omega \text{vec}[E(v_t^2 \sigma_t \sigma_t' + v \lambda_t \sigma_t' + v \sigma_t \lambda_t')].
\] (33)

When \( r = 1 \), by equation (5), we have

\[
\text{vec}[E(\lambda_t \lambda_t')] = \text{vec}[M_1 E(y_t y_t') M_1'] = (M_1 \otimes M_1) \text{vec}[E(y_t y_t')].
\] (34)

Then, by taking vec on both sides of equation (32) and then inserting equations (31) and (34), we have

\[
\Omega \text{vec}[E(v_t^2 \sigma_t \sigma_t' + v \lambda_t \sigma_t' + v \sigma_t \lambda_t')]
\]

\[
= \text{vec}[E(y_t y_t')] - \text{vec}[E(\lambda_t \lambda_t' + v \sigma_t \lambda_t' + v \lambda_t \sigma_t' + v^2 \sigma_t \sigma_t')]
\]

\[
= -\text{vec}[E(v_t^2 \sigma_t \sigma_t' + v \lambda_t \sigma_t' + v \sigma_t \lambda_t')] + \text{vec}[E(y_t y_t')] - \text{vec}[E(\lambda_t \lambda_t')]
\]

\[
= -\text{vec}[E(v_t^2 \sigma_t \sigma_t' + v \lambda_t \sigma_t' + v \sigma_t \lambda_t')] + (I_{N^2} - M_1 \otimes M_1) \text{vec}[E(y_t y_t')].
\]

Therefore, we can write

\[
\text{vec}[E(v_t^2 \sigma_t \sigma_t' + v \lambda_t \sigma_t' + v \sigma_t \lambda_t')] = (\Omega + I_{N^2})^{-1} (I_{N^2} - M_1 \otimes M_1) \text{vec}[E(y_t y_t')].
\] (35)

By multiplying \( \Omega \) on both sides of equation (35) and then inserting equation (33), we obtain

\[
\text{vec}[E(v_t v_t')] = \Omega \text{vec}[E(v_t^2 \sigma_t \sigma_t' + v \lambda_t \sigma_t' + v \sigma_t \lambda_t')]
\]

\[
= \Omega (\Omega + I_{N^2})^{-1} (I_{N^2} - M_1 \otimes M_1) \text{vec}[E(y_t y_t')],
\]

which completes the proof.
Proof for Proposition 4. By taking vec on both sides of equation (18), we have

\[ \operatorname{vec}[E(y, y')] = \sum_{i=0}^{\infty} (\Phi_i \otimes \Phi_i) \operatorname{vec}[E(\nu, \nu')] + \operatorname{vec}(\bar{y}\bar{y}') \].

By applying the result of Lemma 1, we obtain

\[ \operatorname{vec}[E(y, y')] = \Psi_2 \operatorname{vec}[E(y, y')] + \operatorname{vec}(\bar{y}\bar{y}') \],

which can be solved for \( \operatorname{vec}[E(y, y')] \) to obtain equation (19) if and only if all eigenvalues of the matrix \( \Psi_2 \) are less than 1 in modulus.

Proof for Corollary 2. The unconditional first moment can be obtained directly from Proposition 3. For the unconditional second moment, we consider the VMA(\( \infty \)) representation (15) for the \( \text{WAR}(1) \) model. By equation (16), we have

\[ \Phi_i = \mathcal{M}_1 \Phi_{i-1}, \quad (36) \]

with \( \Phi_0 = I_N \). By applying equation (36) recursively, we can see that the \( \text{WAR}(1) \) model can be expressed as a VMA(\( \infty \)) model with parameters \( \Phi_i = \mathcal{M}_1^i (i = 0, 1, \ldots) \). Under the assumption that the second moment exists, we have

\[ \sum_{i=0}^{\infty} (\Phi_i \otimes \Phi_i) = \sum_{i=0}^{\infty} (\mathcal{M}_1^i \otimes \mathcal{M}_1^i) \]

\[ = \sum_{i=0}^{\infty} (\mathcal{M}_1 \otimes \mathcal{M}_1)^i \]

\[ = (I_{N^2} - \mathcal{M}_1 \otimes \mathcal{M}_1)^{-1}. \]
By applying the result of Proposition 4, we obtain

\[
\text{vec}[E(y,y')] = \left[ I_N^2 - (I_N^2 - \mathcal{M}_1 \otimes \mathcal{M}_1)^{-1} \Omega (\Omega + I_N^2)^{-1} \right]^{-1} \text{vec}(\bar{y}\bar{y}')
\]

\[
= (I_N^2 - \mathcal{M}_1 \otimes \mathcal{M}_1)^{-1} \left[ I_N^2 - \Omega (\Omega + I_N^2)^{-1} \right]^{-1} \text{vec}(\bar{y}\bar{y}')
\]

\[
= (I_N^2 - \mathcal{M}_1 \otimes \mathcal{M}_1)^{-1} (\Omega + I_N^2) (\Omega + I_N^2 - \Omega)^{-1} \text{vec}(\bar{y}\bar{y}')
\]

\[
= (I_N^2 - \mathcal{M}_1 \otimes \mathcal{M}_1)^{-1} (\Omega + I_N^2) (I_N^2 - \mathcal{M}_1 \otimes \mathcal{M}_1) \text{vec}(\bar{y}\bar{y}')
\]

which completes the proof.
References


Table 1: Parameter values used in the simulation studies.

<table>
<thead>
<tr>
<th>Param.</th>
<th>Value</th>
<th>Param.</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Model 1: a GCAW(1, 2, 1) model</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$A_1$</td>
<td>0.208</td>
<td>$A_2$</td>
<td>0.073</td>
</tr>
<tr>
<td>0.004</td>
<td>0.022</td>
<td>0.078</td>
<td>0.093</td>
</tr>
<tr>
<td>0.002</td>
<td>0.068</td>
<td>0.012</td>
<td>0.130</td>
</tr>
<tr>
<td>0.042</td>
<td>0.082</td>
<td>0.068</td>
<td>0.009</td>
</tr>
<tr>
<td>0.085</td>
<td>0.048</td>
<td>0.026</td>
<td>0.104</td>
</tr>
<tr>
<td>$B_1$</td>
<td>0.452</td>
<td>$C$</td>
<td>0.041</td>
</tr>
<tr>
<td>0.021</td>
<td>0.917</td>
<td>0.005</td>
<td>0.007</td>
</tr>
<tr>
<td>0.043</td>
<td>0.112</td>
<td>0.001</td>
<td>0.000</td>
</tr>
<tr>
<td>0.529</td>
<td>0.795</td>
<td>0.172</td>
<td>0.075</td>
</tr>
<tr>
<td>$M_1$</td>
<td>0.034</td>
<td>0.795</td>
<td>0.006</td>
</tr>
<tr>
<td>0.005</td>
<td>-0.147</td>
<td>0.146</td>
<td>0.006</td>
</tr>
<tr>
<td>0.536</td>
<td>0.532</td>
<td>0.087</td>
<td>0.075</td>
</tr>
<tr>
<td>$\nu$</td>
<td>15</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

| **Model 2: a CAW(1, 2) (GCAW(1, 2, 0)) model** | | | |
| $A_1$ | 0.216 | $A_2$ | 0.077 |
| 0.007 | 0.098 | 0.055 | 0.172 |
| 0.006 | 0.098 | 0.067 | 0.017 |
| 0.072 | 0.027 | 0.067 | 0.007 |
| 0.619 | 0.146 | 0.075 | 0.000 |
| $B_1$ | 0.462 | $C$ | 0.100 |
| 0.051 | 0.087 | 0.006 | 0.051 |
| 0.072 | 0.312 | 0.067 | 0.000 |
| 0.039 | 0.795 | 0.024 | 0.000 |
| $\nu$ | 15 | | |
Table 2: Summary statistics for the realized variances and covariances obtained by the ARVM estimator.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Mean</th>
<th>Max.</th>
<th>Min.</th>
<th>Std. dev.</th>
<th>Skew.</th>
<th>Kurt.</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Realized variance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AXP ($Y_{11}$)</td>
<td>4.81</td>
<td>248.32</td>
<td>0.08</td>
<td>10.00</td>
<td>9.42</td>
<td>175.71</td>
<td>0.49</td>
</tr>
<tr>
<td>C ($Y_{22}$)</td>
<td>8.23</td>
<td>862.06</td>
<td>0.10</td>
<td>31.83</td>
<td>13.84</td>
<td>284.40</td>
<td>0.35</td>
</tr>
<tr>
<td>GE ($Y_{33}$)</td>
<td>3.48</td>
<td>158.42</td>
<td>0.10</td>
<td>7.90</td>
<td>8.33</td>
<td>110.40</td>
<td>0.51</td>
</tr>
<tr>
<td>HD ($Y_{44}$)</td>
<td>3.59</td>
<td>159.68</td>
<td>0.14</td>
<td>5.84</td>
<td>10.91</td>
<td>238.58</td>
<td>0.47</td>
</tr>
<tr>
<td>IBM ($Y_{55}$)</td>
<td>2.14</td>
<td>69.71</td>
<td>0.08</td>
<td>3.69</td>
<td>7.31</td>
<td>90.06</td>
<td>0.50</td>
</tr>
<tr>
<td><strong>Realized covariance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-AXP ($Y_{21}$)</td>
<td>2.95</td>
<td>146.93</td>
<td>-0.77</td>
<td>8.26</td>
<td>8.80</td>
<td>114.73</td>
<td>0.39</td>
</tr>
<tr>
<td>GE-AXP ($Y_{31}$)</td>
<td>1.93</td>
<td>96.30</td>
<td>-1.37</td>
<td>4.69</td>
<td>8.07</td>
<td>108.76</td>
<td>0.46</td>
</tr>
<tr>
<td>HD-AXP ($Y_{41}$)</td>
<td>1.77</td>
<td>83.32</td>
<td>-1.68</td>
<td>3.89</td>
<td>8.11</td>
<td>115.58</td>
<td>0.49</td>
</tr>
<tr>
<td>IBM-AXP ($Y_{51}$)</td>
<td>1.33</td>
<td>42.29</td>
<td>-1.71</td>
<td>2.94</td>
<td>6.88</td>
<td>67.63</td>
<td>0.46</td>
</tr>
<tr>
<td>GE-C ($Y_{32}$)</td>
<td>2.30</td>
<td>136.74</td>
<td>-0.56</td>
<td>6.67</td>
<td>10.11</td>
<td>150.74</td>
<td>0.42</td>
</tr>
<tr>
<td>HD-C ($Y_{42}$)</td>
<td>2.04</td>
<td>105.59</td>
<td>-0.24</td>
<td>5.13</td>
<td>9.50</td>
<td>145.01</td>
<td>0.48</td>
</tr>
<tr>
<td>IBM-C ($Y_{52}$)</td>
<td>1.53</td>
<td>87.09</td>
<td>-2.06</td>
<td>3.95</td>
<td>10.06</td>
<td>157.43</td>
<td>0.43</td>
</tr>
<tr>
<td>HD-GE ($Y_{43}$)</td>
<td>1.53</td>
<td>81.77</td>
<td>-0.29</td>
<td>3.41</td>
<td>9.47</td>
<td>162.45</td>
<td>0.49</td>
</tr>
<tr>
<td>IBM-GE ($Y_{53}$)</td>
<td>1.23</td>
<td>48.33</td>
<td>-0.23</td>
<td>2.66</td>
<td>7.62</td>
<td>90.81</td>
<td>0.47</td>
</tr>
<tr>
<td>IBM-HD ($Y_{54}$)</td>
<td>1.19</td>
<td>74.84</td>
<td>-0.60</td>
<td>2.80</td>
<td>11.84</td>
<td>240.75</td>
<td>0.49</td>
</tr>
</tbody>
</table>

Note: $d$ is the log-periodogram regression estimate of the fractional integration parameter based on the $m = T^{4/5} = 485$ lowest-frequency periodogram ordinates. The asymptotic standard error for all of the $d$ estimates is $\pi \times (24m)^{-1/2} = 0.0291$. 

**ACCEPTED MANUSCRIPT**
Table 3: Summary statistics for the realized variances and covariances obtained by the multivariate realized kernel estimator.

<table>
<thead>
<tr>
<th>Stock</th>
<th>Mean</th>
<th>Max.</th>
<th>Min.</th>
<th>Std. dev.</th>
<th>Skew.</th>
<th>Kurt.</th>
<th>$d$</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Realized variance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>AXP ($Y_{11}$)</td>
<td>5.70</td>
<td>324.95</td>
<td>0.01</td>
<td>12.69</td>
<td>11.94</td>
<td>203.42</td>
<td>0.55</td>
</tr>
<tr>
<td>C ($Y_{22}$)</td>
<td>8.02</td>
<td>980.10</td>
<td>0.08</td>
<td>40.96</td>
<td>19.05</td>
<td>271.97</td>
<td>0.38</td>
</tr>
<tr>
<td>GE ($Y_{33}$)</td>
<td>4.06</td>
<td>193.53</td>
<td>0.10</td>
<td>9.59</td>
<td>12.09</td>
<td>121.83</td>
<td>0.44</td>
</tr>
<tr>
<td>HD ($Y_{44}$)</td>
<td>3.80</td>
<td>135.29</td>
<td>0.15</td>
<td>5.35</td>
<td>12.90</td>
<td>252.63</td>
<td>0.45</td>
</tr>
<tr>
<td>IBM ($Y_{55}$)</td>
<td>2.38</td>
<td>78.55</td>
<td>0.02</td>
<td>4.39</td>
<td>7.99</td>
<td>102.52</td>
<td>0.49</td>
</tr>
<tr>
<td><strong>Realized covariance</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C-AXP ($Y_{21}$)</td>
<td>2.25</td>
<td>119.22</td>
<td>-1.05</td>
<td>6.72</td>
<td>5.34</td>
<td>84.20</td>
<td>0.33</td>
</tr>
<tr>
<td>GE-AXP ($Y_{31}$)</td>
<td>1.69</td>
<td>133.27</td>
<td>-2.91</td>
<td>5.53</td>
<td>14.79</td>
<td>120.58</td>
<td>0.49</td>
</tr>
<tr>
<td>HD-AXP ($Y_{41}$)</td>
<td>1.50</td>
<td>95.55</td>
<td>-1.03</td>
<td>3.73</td>
<td>7.13</td>
<td>95.40</td>
<td>0.39</td>
</tr>
<tr>
<td>IBM-AXP ($Y_{51}$)</td>
<td>1.24</td>
<td>62.69</td>
<td>-1.39</td>
<td>3.50</td>
<td>10.61</td>
<td>60.68</td>
<td>0.42</td>
</tr>
<tr>
<td>GE-C ($Y_{32}$)</td>
<td>2.51</td>
<td>116.42</td>
<td>-0.36</td>
<td>5.36</td>
<td>6.20</td>
<td>132.91</td>
<td>0.44</td>
</tr>
<tr>
<td>HD-C ($Y_{42}$)</td>
<td>1.83</td>
<td>97.95</td>
<td>-0.10</td>
<td>5.09</td>
<td>8.54</td>
<td>147.20</td>
<td>0.39</td>
</tr>
<tr>
<td>IBM-C ($Y_{52}$)</td>
<td>1.19</td>
<td>67.05</td>
<td>-3.38</td>
<td>3.71</td>
<td>9.27</td>
<td>178.10</td>
<td>0.49</td>
</tr>
<tr>
<td>HD-GE ($Y_{43}$)</td>
<td>1.22</td>
<td>74.28</td>
<td>-0.44</td>
<td>2.58</td>
<td>6.51</td>
<td>146.95</td>
<td>0.46</td>
</tr>
<tr>
<td>IBM-GE ($Y_{53}$)</td>
<td>1.01</td>
<td>42.90</td>
<td>-0.81</td>
<td>2.37</td>
<td>6.96</td>
<td>85.08</td>
<td>0.47</td>
</tr>
<tr>
<td>IBM-HD ($Y_{54}$)</td>
<td>1.01</td>
<td>57.27</td>
<td>-0.32</td>
<td>2.38</td>
<td>9.42</td>
<td>193.31</td>
<td>0.47</td>
</tr>
</tbody>
</table>

Note: $d$ is the log-periodogram regression estimate of the fractional integration parameter based on the $m = T^{4/5} = 485$ lowest-frequency periodogram ordinates. The asymptotic standard error for all of the $d$ estimates is $\pi \times (24m)^{-1/2} = 0.0291$. 


Table 4: Goodness of fit measures for the GCAW models.

<table>
<thead>
<tr>
<th>(p, q, r)</th>
<th>dim(θ)</th>
<th>$\ell_p(\hat{\theta})$</th>
<th>BIC</th>
<th>meig1</th>
<th>meig2</th>
<th>p-value for the portmanteau test using 100 lags of the standardized residuals $c_{it}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCAW (0, 0)</td>
<td>16</td>
<td>0.5074</td>
<td>101228</td>
<td>0</td>
<td>0.256</td>
<td>0.00</td>
</tr>
<tr>
<td>GCAW (0, 0, 2)</td>
<td>41</td>
<td>-46044</td>
<td>92805</td>
<td>0.276</td>
<td>0.256</td>
<td>0.00</td>
</tr>
<tr>
<td>GCAW (0, 0, 3)</td>
<td>66</td>
<td>-45923</td>
<td>92557</td>
<td>0.281</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>GCAW (p, q, r)</td>
<td>91</td>
<td>-45880</td>
<td>92464</td>
<td>0.284</td>
<td>0.00</td>
<td>0.00</td>
</tr>
<tr>
<td>GCAW (p, q)</td>
<td>141</td>
<td>-13835</td>
<td>28807</td>
<td>0.994</td>
<td>1.103</td>
<td>0.23</td>
</tr>
</tbody>
</table>

Note: dim(θ) is the number of parameters, $\ell_p(\hat{\theta})$ is the optimized log-likelihood. meig1 and meig2 are the maximum eigenvalues of the matrices $\Psi_1$ and $\Psi_2$ given in equations (14) and (20), respectively. Bold p-value indicates the serial correlation in the corresponding standardized residual is significant at the 1% significance level.
Table 5: ML-estimates for the GCAW(2, 2, 0) and GCAW(2, 2, 1) models.

<table>
<thead>
<tr>
<th>Param.</th>
<th>Estimate</th>
<th>Param.</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCAW(2, 2, 0) (CAW(2, 2))</td>
<td></td>
<td>GCAW(2, 2, 1)</td>
<td></td>
</tr>
<tr>
<td>$A_1$</td>
<td>0.142</td>
<td>0.012</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>0.171</td>
<td>−0.012</td>
</tr>
<tr>
<td></td>
<td>0.006</td>
<td>0.010</td>
<td>0.139</td>
</tr>
<tr>
<td></td>
<td>0.011</td>
<td>0.006</td>
<td>−0.003</td>
</tr>
<tr>
<td></td>
<td>0.003</td>
<td>0.003</td>
<td>−0.001</td>
</tr>
<tr>
<td>$B_1$</td>
<td>0.726</td>
<td>−0.069</td>
<td>−0.142</td>
</tr>
<tr>
<td></td>
<td>0.027</td>
<td>0.376</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>0.012</td>
<td>−0.113</td>
<td>0.619</td>
</tr>
<tr>
<td></td>
<td>−0.003</td>
<td>−0.033</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>0.009</td>
<td>−0.010</td>
<td>0.026</td>
</tr>
<tr>
<td>$C$</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>−0.002</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.018</td>
<td>0.023</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>0.048</td>
<td>−0.043</td>
<td>−0.012</td>
</tr>
<tr>
<td></td>
<td>−0.002</td>
<td>0.004</td>
<td>−0.032</td>
</tr>
<tr>
<td>$\nu$</td>
<td></td>
<td>20.685</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Param.</th>
<th>Estimate</th>
<th>Param.</th>
<th>Estimate</th>
</tr>
</thead>
<tbody>
<tr>
<td>GCAW(2, 2, 1)</td>
<td></td>
<td></td>
<td>M1</td>
</tr>
<tr>
<td>$A_1$</td>
<td>0.137</td>
<td>0.008</td>
<td>0.007</td>
</tr>
<tr>
<td></td>
<td>0.009</td>
<td>0.163</td>
<td>−0.012</td>
</tr>
<tr>
<td></td>
<td>0.016</td>
<td>0.000</td>
<td>0.136</td>
</tr>
<tr>
<td></td>
<td>0.015</td>
<td>0.004</td>
<td>−0.005</td>
</tr>
<tr>
<td></td>
<td>0.005</td>
<td>0.003</td>
<td>−0.001</td>
</tr>
<tr>
<td>$B_1$</td>
<td>0.734</td>
<td>−0.046</td>
<td>−0.131</td>
</tr>
<tr>
<td></td>
<td>0.010</td>
<td>0.427</td>
<td>−0.021</td>
</tr>
<tr>
<td></td>
<td>−0.004</td>
<td>−0.031</td>
<td>0.582</td>
</tr>
<tr>
<td></td>
<td>−0.004</td>
<td>−0.018</td>
<td>0.070</td>
</tr>
<tr>
<td></td>
<td>0.006</td>
<td>−0.010</td>
<td>0.022</td>
</tr>
<tr>
<td>$C$</td>
<td>0.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>−0.005</td>
<td>0.000</td>
<td></td>
</tr>
<tr>
<td></td>
<td>−0.023</td>
<td>0.022</td>
<td>0.000</td>
</tr>
<tr>
<td></td>
<td>−0.038</td>
<td>−0.036</td>
<td>−0.036</td>
</tr>
<tr>
<td></td>
<td>−0.004</td>
<td>0.003</td>
<td>0.006</td>
</tr>
<tr>
<td>$\nu$</td>
<td></td>
<td>20.778</td>
<td></td>
</tr>
</tbody>
</table>
Table 6: Likelihood ratio test results.

<table>
<thead>
<tr>
<th>Order</th>
<th>$H_0: \Lambda_t = 0$</th>
<th>$H_0: \Sigma_t = CC'$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>LR statistic</td>
<td>Critical value</td>
</tr>
<tr>
<td>(0, 1, 1)</td>
<td>2760.60</td>
<td>37.65</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>1055.53</td>
<td>37.65</td>
</tr>
<tr>
<td>(1, 2, 1)</td>
<td>422.57</td>
<td>37.65</td>
</tr>
<tr>
<td>(2, 1, 1)</td>
<td>751.54</td>
<td>37.65</td>
</tr>
<tr>
<td>(2, 2, 1)</td>
<td>343.72</td>
<td>37.65</td>
</tr>
<tr>
<td>(2, 3, 1)</td>
<td>296.39</td>
<td>37.65</td>
</tr>
<tr>
<td>(3, 2, 1)</td>
<td>300.60</td>
<td>37.65</td>
</tr>
<tr>
<td>(3, 3, 1)</td>
<td>284.15</td>
<td>37.65</td>
</tr>
<tr>
<td>(0, 1, 2)</td>
<td>3572.63</td>
<td>67.50</td>
</tr>
<tr>
<td>(1, 1, 2)</td>
<td>1076.40</td>
<td>67.50</td>
</tr>
<tr>
<td>(1, 2, 2)</td>
<td>460.32</td>
<td>67.50</td>
</tr>
<tr>
<td>(2, 1, 2)</td>
<td>863.50</td>
<td>67.50</td>
</tr>
<tr>
<td>(2, 2, 2)</td>
<td>378.31</td>
<td>67.50</td>
</tr>
<tr>
<td>(2, 3, 2)</td>
<td>330.74</td>
<td>67.50</td>
</tr>
<tr>
<td>(3, 2, 2)</td>
<td>337.77</td>
<td>67.50</td>
</tr>
<tr>
<td>(3, 3, 2)</td>
<td>314.28</td>
<td>67.50</td>
</tr>
</tbody>
</table>

Note: $\Lambda_t = 0$ implies $M_1 = \cdots = M_r = 0$ and $\Sigma_t = CC'$ implies $A_1 = \cdots = A_q = B_1 = \cdots = B_p = 0$. 

Accepted Manuscript
Table 7: Forecasting accuracy for the GCAW models.

<table>
<thead>
<tr>
<th>(p, q, r)</th>
<th>dim(θ)</th>
<th>Normal period</th>
<th>Crisis period</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>h = 1</td>
<td>h = 5</td>
</tr>
<tr>
<td>GCAW(0, 0, 0) (central Wishart distribution)</td>
<td>16</td>
<td>8.9537</td>
<td>8.9755</td>
</tr>
<tr>
<td>GCAW(0, 0, r) (WAR(r))</td>
<td>41</td>
<td>7.2278</td>
<td>8.9539</td>
</tr>
<tr>
<td>(0, 0, 1)</td>
<td>66</td>
<td>7.2736</td>
<td>8.9488</td>
</tr>
<tr>
<td>(0, 0, 2)</td>
<td>91</td>
<td>7.2775</td>
<td>8.9380</td>
</tr>
<tr>
<td>(0, 0, 3)</td>
<td>41</td>
<td>6.0821</td>
<td>7.9238</td>
</tr>
<tr>
<td>(0, 1, 1)</td>
<td>66</td>
<td>5.9534</td>
<td>7.5467</td>
</tr>
<tr>
<td>(1, 1, 0)</td>
<td>91</td>
<td>5.9600</td>
<td>7.5196</td>
</tr>
<tr>
<td>(1, 1, 1)</td>
<td>116</td>
<td>9.8737</td>
<td>7.4524</td>
</tr>
<tr>
<td>(1, 2, 0)</td>
<td>141</td>
<td>5.8638</td>
<td>7.2894</td>
</tr>
<tr>
<td>(2, 1, 1)</td>
<td>91</td>
<td>5.9347</td>
<td>7.5467</td>
</tr>
<tr>
<td>(2, 1, 0)</td>
<td>66</td>
<td>5.9279</td>
<td>7.5196</td>
</tr>
<tr>
<td>(2, 2, 0)</td>
<td>91</td>
<td>5.8703</td>
<td>7.3177</td>
</tr>
<tr>
<td>(3, 2, 0)</td>
<td>166</td>
<td>5.7132</td>
<td>7.3593</td>
</tr>
<tr>
<td>(3, 3, 0)</td>
<td>66</td>
<td>5.9977</td>
<td>7.4556</td>
</tr>
<tr>
<td>(0, 1, 2)</td>
<td>116</td>
<td>5.8063</td>
<td>7.2183</td>
</tr>
<tr>
<td>(1, 2, 1)</td>
<td>141</td>
<td>5.7773</td>
<td>7.2500</td>
</tr>
<tr>
<td>(2, 2, 1)</td>
<td>91</td>
<td>5.7988</td>
<td>7.2296</td>
</tr>
<tr>
<td>(2, 3, 1)</td>
<td>141</td>
<td>5.7126</td>
<td>7.1626</td>
</tr>
<tr>
<td>(3, 2, 1)</td>
<td>166</td>
<td>5.7301</td>
<td>7.2242</td>
</tr>
<tr>
<td>(3, 3, 1)</td>
<td>191</td>
<td>5.6819</td>
<td>7.2089</td>
</tr>
<tr>
<td>(0, 1, 3)</td>
<td>91</td>
<td>5.9807</td>
<td>7.4336</td>
</tr>
<tr>
<td>(1, 2, 1)</td>
<td>116</td>
<td>5.8606</td>
<td>7.3070</td>
</tr>
<tr>
<td>(2, 2, 1)</td>
<td>141</td>
<td>5.7647</td>
<td>7.0782</td>
</tr>
<tr>
<td>(2, 3, 1)</td>
<td>141</td>
<td>5.7315</td>
<td>7.0972</td>
</tr>
<tr>
<td>(3, 2, 1)</td>
<td>166</td>
<td>5.6874</td>
<td>7.0330</td>
</tr>
<tr>
<td>(3, 3, 1)</td>
<td>191</td>
<td>5.6932</td>
<td>7.0533</td>
</tr>
<tr>
<td>(0, 1, 2)</td>
<td>216</td>
<td>5.6818</td>
<td>7.0575</td>
</tr>
</tbody>
</table>

Note: dim(θ) is the number of parameters. FNₙ given in equation (26) is the average forecast error (for 252 forecasts) during the normal period (from 2 July 2007 to 30 June 2008) and crisis period (from 1 July 2008 to 30 June 2009). Bold number indicates the smallest value of the average forecast error.
Figure 1: Time series of the realized variances and covariances.
Figure 2: Fitted values of the realized variances and covariances under the GCAW(2, 2, 1) model; the solid and dotted lines indicate the observed and fitted values, respectively.