Regensburger
DISKUSSIONSBEITRÄGE
zur Wirtschaftswissenschaft

University of Regensburg Working Papers in Business, Economics and Management Information Systems

Matrix Box-Cox Models for Multivariate Realized Volatility

Roland Weigand*

March 21, 2014

Nr. 478

JEL Classification: C14, C32, C51, C53, C58

Key Words: Realized covariance matrix, dynamic correlation, semiparametric estimation, density forecasting

* Roland Weigand is a researcher at the Institute of Employment Research, 90478 Nuremberg, Germany, and doctoral student at the Department of Economics and Econometrics, Faculty of Business, Economics and Management Information Systems at the University of Regensburg, 93040 Regensburg, Germany
Phone: +49-911-179-3291, E-mail: roland.weigand[at]iab.de

Matrix Box-Cox Models for Multivariate Realized Volatility

Roland Weigand*

Institute for Employment Research (IAB), D-90478 Nuremberg

March 2014

Abstract. We propose flexible models for multivariate realized volatility dynamics which involve generalizations of the Box-Cox transform to the matrix case. The matrix Box-Cox model of realized covariances (MBC-RCov) is based on transformations of the covariance matrix eigenvalues, while for the Box-Cox dynamic correlation (BC-DC) specification the variances are transformed individually and modeled jointly with the correlations. We estimate transformation parameters by a new multivariate semiparametric estimator and discuss bias-corrected point and density forecasting by simulation. The methods are applied to stock market data where excellent in-sample and out-of-sample performance is found.

Keywords. Realized covariance matrix, dynamic correlation, semiparametric estimation, density forecasting.

JEL-Classification. C14, C32, C51, C53, C58.

*Phone: +49 911 179 3291, E-Mail: roland.weigand@iab.de

This research has partly been done at the Institute of Economics and Econometrics of the University of Regensburg. The author is very grateful for helpful comments by Rolf Tschernig, Enzo Weber and participants of the Humboldt–Copenhagen Conference on Financial Econometrics 2013. Support by BayEFG is gratefully acknowledged.

1 Introduction

Dynamic modeling of multivariate financial volatility has recently gained significant interest. On the one hand, it constitutes an essential part of portfolio decisions, in empirical asset pricing models and for derivative analysis. On the other hand, recent financial crises have accentuated the importance of quantifying systemic risk. The latter also requires multivariate rather than univariate models. Such models require a precise measure of the otherwise latent asset variance and covariance processes and a framework for modeling the dynamics. Precise measures are available due to recent and significant achievements on multivariate realized financial volatility modelling; see, e.g., Andersen et al. (2003) and Barndorff-Nielsen and Shephard (2004).

There are several approaches for modeling covariance dynamics. A prominent model class is based on conditionally Wishart distributed processes (see, e.g., Golosnoy et al.; 2012). Alternatively, linear vector time series models are applied to specific transformations of realized covariance matrices. The latter approach has the advantage of simplicity; model estimation, checking and inference is implemented in econometric software packages, while suitable ways of handling high-dimensional panels of time series are well-established. Various transformations have recently been suggested: Chiriac and Voev (2011), for instance, use the elements of a triangular matrix square-root transform, while the matrix logarithm has been considered by Bauer and Vorkink (2011) as well as Gribisch (2013). For these models, fitted covariance matrices and out-of-sample forecasts are automatically positive definite through the corresponding retransformations.

Likewise, approaches that separate variance and correlation dynamics, so called dynamic correlation (DC) models have been a fruitful direction of research. With appropriate factor or panel structure assumptions, Golosnoy and Herwartz (2012) model the z-transformed realized correlations (cf. (6) below). Correlation eigenvalues along with locally constant eigenvectors, sampled at different frequencies, are used by Hautsch et al. (2014). Separate realized variance and correlation dynamics in mixed frequency models are also investigated by Halbleib and Voev (2011).

In the univariate time series literature, where transformation-based methods have a long tradition, the model of Box and Cox (1964) has become popular to find a suitable transform prior to ARIMA analysis (Box and Jenkins; 1970). Similar approaches were used for univariate volatility modeling, e.g., by Higgins and Bera (1992), Yu et al. (2006), Zhang and King (2008) and Goncalves and Meddahi (2011).

We propose two flexible models in the spirit of Box and Cox (1964) for the dynamic multivariate realized volatility setup. Both generalize the univariate Box-Cox transform to the matrix case and contain several well-known transforms as special cases. The matrix Box-Cox model of realized covariances (MBC-RCov) is based on transformations of the covariance matrix eigenvalues. On the other hand, for the Box-Cox dynamic correlation (BC-DC) model, the variances are transformed individually and modeled together with the z-transformed correlations.

We introduce a semiparametric estimator of the transformation parameters in the multivariate setup by generalizing the univariate approach of Proietti and Lütkepohl (2013). It does not require the specification of a dynamic model and makes a computationally simple two-step approach feasible. A simulation-based forecasting procedure is presented to reduce the bias of the naïve re-transform forecasts. Simulated paths of the realized volatilities may also be used to obtain density forecasts of the daily returns which will often be the aim of studying covariance matrix dynamics.

We apply these methods to the data set of Chiriac and Voev (2011) and find that a sparse vector autoregressive vector moving average (VARMA) specification provides a reasonable fit to the transformed series. A pseudo out-of-sample forecast comparison is conducted, where the BC-DC specification either with estimated transformation parameters or restricted to the logarithmic case emerges as favorable in practice. Bias correction provides significant improvements over the naïve forecasts. Notably, also the conditional Wishart models as popular benchmarks are outperformed by our transformation-based approach. These results are robust to different dynamic specifications and remain qualitatively intact for most of the loss functions recently used for evaluations of this kind.

The paper is organized as follows: In section 2 the new models are introduced. Parameter estimation and forecasting is described in sections 3 and 4, respectively. Section 5 presents the estimation results, while section 6 contains the out-of-sample forecast evaluation. Section 7 concludes.

2 Multivariate Box-Cox volatility models

In univariate regression and time series models, the Box-Cox transformation (Box and Cox; 1964) has been applied to obtain a linear, homoscedastic specification for the transformed dependent variable. For a scalar $x > 0$, it is parameterized by δ and given by

$$h(x;\delta) = \begin{cases} \frac{x^{\delta}-1}{\delta} & \text{for } \delta \neq 0, \\ \log(x) & \text{for } \delta = 0. \end{cases} \tag{1}$$

For specific choices of δ, the transform corresponds to a linear mapping of the raw series ($\delta = 1$) or of various popular transforms such as the square root ($\delta = 0.5$), the logarithm ($\delta = 0$) and the inverse ($\delta = -1$). The reverse transform is given by

$$h^{-1}(y;\delta) = \begin{cases} (\delta y + 1)^{\frac{1}{\delta}} & \text{for } \delta \neq 0, \\ \exp(y) & \text{for } \delta = 0, \end{cases} \tag{2}$$

which is defined for $y > -\frac{1}{\delta}$ if $\delta > 0$ and for $y < -\frac{1}{\delta}$ if $\delta < 0$, and gives strictly positive values.

2.1 The matrix Box-Cox model of realized covariances

To generalize the Box-Cox method for modeling covariance matrices we define a matrix version of the latter, the matrix Box-Cox (MBC) transform. For a positive definite $(k \times k)$ covariance matrix X_t, and $t = 1, 2, \ldots$ denoting time periods, we suggest to apply Box-Cox transformations to the eigenvalues of X_t, each with a distinct transformation parameter collected in $\delta = (\delta_1, \ldots, \delta_k)'$,

$$Y_t(\delta) := H(X_t;\delta) = V_t \begin{pmatrix} h(\lambda_{1t};\delta_1) & 0 & \cdots & 0 \\ 0 & h(\lambda_{2t};\delta_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & h(\lambda_{kt};\delta_k) \end{pmatrix} V_t'. \tag{3}$$

Here $\lambda_{1t} \geq \ldots \geq \lambda_{kt} \geq 0$ are the eigenvalues of X_t, $h(\lambda_{it};\delta_i)$, $i = 1, \ldots, k$, are their univariate Box-Cox transforms and V_t denotes the matrix of eigenvectors of X_t.

To understand the consequences of the MBC approach for modelling covariance matrices, it is useful to consider the inverse transformation, applied to a symmetric $(k \times k)$ matrix Y_t,

$$H^{-1}(Y_t;\delta) = V_t \begin{pmatrix} h^{-1}(\lambda^y_{1t};\delta_1) & 0 & \dots & 0 \\ 0 & h^{-1}(\lambda^y_{2t};\delta_2) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & h^{-1}(\lambda^y_{kt};\delta_k) \end{pmatrix} V_t'. \tag{4}$$

Here, by λ^y_{jt}, $j = 1,\dots,k$, we denote the eigenvalues of Y_t, while V_t contains the eigenvectors of both Y_t and $H^{-1}(Y_t;\delta)$, which remain unaffected by the transform. Notably, when the inverse MBC transform is well defined and applied to a symmetric matrix, the re-transformed fitted or forecasted matrices are always positive definite.

The reverse Box-Cox transform is not always well-defined, however. As mentioned below (2), existence of h^{-1} and hence of H^{-1} requires that the eigenvalues satisfy certain restrictions, namely $\lambda^y_{jt}(\delta_j) > -\frac{1}{\delta_j}$ for $\delta_j > 0$ and $\lambda^y_{jt}(\delta_j) < -\frac{1}{\delta_j}$ for $\delta_j < 0$. This requirement limits the set of feasible values of δ for a given sequence of matrices (e.g., forecasts) to which the inverse transform has to be applied. Our empirical results for stock market data suggest that this potential drawback may be irrelevant as long as the applied transformation parameters are not chosen grossly at odds with estimates from the data (i.e. for $\delta_j > -0.25$ in our application).

As in the univariate setup, the matrix transform contains as special cases linear combinations of the raw matrix entries ($\delta_1 = \delta_2 = \ldots = \delta_k = 1$), of the (symmetric) matrix square root $\delta_1 = \delta_2 = \ldots = \delta_k = 0.5$), of the matrix logarithm ($\delta_1 = \delta_2 = \ldots = \delta_k = 0$) and of the inverse ($\delta_1 = \delta_2 = \ldots = \delta_k = -1$). It thus incorporates several empirically relevant approaches to covariance modeling within a common framework. We call this approach for modeling and forecasting multivariate realized volatility the matrix Box-Cox model of realized covariances (MBC-RCov).

For all periods $t = 1,\dots,T$, the MBC-transform is applied to the realized covariance matrices X_t for an appropriate vector of parameters δ. In this way we obtain a sequence of symmetric matrices $Y_t(\delta)$ from which only the lower triangular elements (including the main diagonal) need to be modeled. A time series model is thus fitted only to the $k(k+1)/2$-

dimensional vector process $y_t(\delta) := \text{vech}(Y_t(\delta))^1$. For generality, we assume a linear process

$$y_t(\delta) = \sum_{j=0}^{\infty} \Psi_j(\theta) u_{t-j}, \qquad u_t \sim IID(0; \Sigma_u), \tag{5}$$

with $\Psi_0 = I$. We let θ as well as Σ_u consist of unknown parameters. Specific models will be considered in the empirical application in section 5. Here, we apply diagonal vector autoregressive moving average (VARMA) models as well as fractionally integrated VARMA (VARFIMA) and multivariate heterogeneous autoregressive (HAR) models.

2.2 The Box-Cox dynamic correlation model

As an alternative to the matrix version of the Box-Cox transform, we consider a decomposition of variances and correlations. Applying the Box-Cox transform to the individual asset variances we introduce the Box-Cox dynamic correlation (BC-DC) model. In the spirit of dynamic conditional correlation models (Engle; 2002), we write $X_t = D_t R_t D_t$, where $D_t = \text{diag}(\sqrt{X_{11,t}}, \ldots, \sqrt{X_{kk,t}})$ is a diagonal matrix containing the univariate realized standard deviations while R_t is the sequence of realized correlation matrices. Applying Fisher's z-transformation

$$\tilde{R}_{ij,t} := \frac{1}{2} \log \frac{1 + R_{ij,t}}{1 - R_{ij,t}} \tag{6}$$

to the correlations has several advantages as compared to using the raw correlations (see Golosnoy and Herwartz; 2012), so that we propose modelling the vector time series

$$z_t(\delta) := g(X_t; \delta) := (h(X_{11,t}; \delta_1), \ldots, h(X_{kk,t}; \delta_k), \tilde{R}_{21,t}, \tilde{R}_{31,t}, \ldots, \tilde{R}_{k\,k-1,t})', \tag{7}$$

as a linear process analogous to (5).

The inverse BC-DC transform g^{-1}, when applied to forecasted z_{T+h}, yields positive variances due to the inverse Box-Cox and correlations in the range $(-1; 1)$ due to the inverse Fisher transform. In contrast to the matrix Box-Cox approach, positive definiteness is not guaranteed for $k > 2$, however.[2] Whenever positive definiteness fails, it has to be enforced and a well-

[1] A different strategy would be to fit a model directly to the transformed eigenvalues and free elements of the eigenvectors, analogously to the approach of Hautsch et al. (2014). They find that the eigenvectors are rather noisy and unstable at daily frequency which is not the case for our vech-transformation.

[2] As a counterexample where the unrestricted forecasts do not yield a valid correlation matrix, consider $k = 3$ and suppose that the inverse Fisher transform gives $R_{12,t} = R_{13,t} = 0.8$ along with $R_{23,t} = -0.8$. The quadratic form $\gamma' R_t \gamma$ is negative, e.g., for $\gamma = (1, -1, -1)'$.

conditioned matrix must be obtained by some sort of eigenvalue trimming or shrinkage procedure. Positive definiteness, however, is not problematic empirically even in high-dimensional stock market applications for z-transformed correlation matrices as the results of Golosnoy and Herwartz (2012) suggest. Compared to the MBC-RCov approach, the estimated dynamics of the linear model (5) fitted to $z_t(\delta)$ are easily interpreted. The matrix Σ_u, for example, provides guidance about the extent of instantaneous co-movement within groups of variances or correlations but also between correlations and variances. Dynamic spill-overs may be modeled by non-diagonal specifications for $\Psi_j(\theta)$.

In addition to enabling a linear homoskedastic specification, the Box-Cox transform has originally been introduced to reduce the deviation from normality of the involved variables or model residuals. However, for the univariate transform (1) it holds that $h(x;\delta) > -\frac{1}{\delta}$ for $\delta > 0$ and $h(x;\delta) < -\frac{1}{\delta}$ for $\delta < 0$. Due to its bounded support, hence, the BC-transformed variable cannot literally be Gaussian whenever $\delta \neq 0$; see, e.g., Amemiya and Powell (1981). Merits of the transform even in cases where Gaussianity fails have been pointed out by Draper and Cox (1969). Although in the matrix case the MBC-transformed series are not individually bounded, the same logic implies that the MBC- and BD-DC-transformed series cannot be exactly multivariate normal. We do not need the Gaussianity assumption at this stage but empirically assess whether the transformed data are at least *approximately* Gaussian later on.

3 Semiparametric estimation of the transformation parameter

In this section we discuss semiparametric estimation of the vector of transformation parameters δ. Among others, Han (1987) has proposed a semiparametric approach to estimate the transformation parameter of a single variable. Likewise, the recently developed estimator of Proietti and Lütkepohl (2013) for time series data does not involve specifying a parametric dynamic model. It is computed by minimizing a frequency-domain estimate of the prediction error variance of the transformed series. In the following, we generalize their approach to multivariate BC-DC and MBC-RCov setups. For the BC-DC model, our multivariate method provides a potentially more efficient estimator than applying the univariate estimator to all k variance series individually and allows to impose cross-equation restrictions. Moreover, in the MBC-RCov context, estimation is inherently multivariate and hence the existing semiparametric approaches would not be applicable without modifications.

In multivariate (vector) Box-Cox regression models, where each of the k nonnegative endogenous variables, say realized variances $(X_{11,t}, \ldots, X_{kk,t})'$, are transformed individually, the standard estimation strategy has been maximum likelihood under the auxiliary assumption of Gaussian transformed variables; see, e.g., Velilla (1993). Maximum likelihood estimation can be straightforwardly extended to the MBC-RCov model, as we outline in Appendix A. In case of dynamic models, the likelihood is simultaneously maximized with respect to both, the dynamic and the transformation parameters. In contrast, our approach allows the researcher to proceed in two steps: After the estimation of the transformation parameters, which involves a k-dimensional optimization for both the BC-DC and MBC-RCov approach, the dynamic model specification and estimation is carried out for the transformed series as if δ was known.

To sketch our semiparametric approach for a generic k-dimensional vector process x_t with strictly positive elements, we consider the Jacobian of the vector Box-Cox transform

$$J_t(\delta) := \left.\frac{\partial(h(x_1;\delta_1), \ldots, h(x_k;\delta_k))'}{\partial x'}\right|_{x=x_t},$$

such that a normalized transform with unit Jacobi determinant is given by

$$\xi_t(\delta) := \left|\prod_{s=1}^{T} J_s(\delta)\right|^{-\frac{1}{kT}} (h(x_{1,t};\delta_1), \ldots, h(x_{k,t};\delta_k))'. \tag{8}$$

The transformed values $h(x_{j,t};\delta_j)$ are corrected for the change in scale induced by the Box-Cox function. The Jacobian is diagonal with elements given by $J_{jj,s}(\delta) = x_{j,s}^{\delta_j - 1}$, so that $\xi_t(\delta)$ is easily computed from $\{x_t\}_{t=1}^{T}$ and δ. Alternatively, the well-known normalization $\check{\xi}_{j,t}(\delta_j) := (\prod_{s=1}^{T} x_{j,s})^{\frac{\delta_j - 1}{T}} h(x_{j,t};\delta_j)$ can be applied to the individual time series. It also succeeds in obtaining scale-invariance and gives numerically identical results for the estimation procedure described in this section.

Without referring to a specific dynamic process, denote the one-step ahead prediction error as $\eta_t(\delta) := \xi_t(\delta) - Proj(\xi_t(\delta)|\mathcal{I}_{t-1})$, where $Proj(\cdot|\mathcal{I}_{t-1})$ is the best linear predictor of a time series given an information set $\mathcal{I}_{t-1}$ which consists of the series' own past in this case and let $\Sigma_\eta(\delta) := Var(\eta_t(\delta))$. Under the assumption that there exists a vector δ^* for which $E(\xi_t(\delta^*)|\mathcal{I}_{t-1})$ is linear in $\xi_{t-j}(\delta^*)$, $j \geq 1$, we characterize this true value δ^* as minimizing the determinant of the prediction error covariance matrix $|\Sigma_\eta(\delta)|$, the so-called generalized variance of $\eta_t(\delta)$.

A *least generalized variance* estimator for δ becomes feasible by utilizing the nonparametric methods proposed by Jones (1976) and further developed by Mohanty and Pourahmadi (1996) to obtain nonparametric estimates of $|\Sigma_\eta(\delta)|$, for a given δ. This generalized prediction error variance is related to the $(k \times k)$ spectral density matrix $g_\xi(\omega)$ of ξ_t by a multivariate extension of the Szegö-Kolmogoroff-Formula (cf. Priestley; 1982, p. 761),

$$\log|\Sigma_\eta| = \frac{1}{2\pi}\int_{-\pi}^{\pi} \log 2\pi |g_\xi(\omega)| d\omega.$$

In practice, the integral may be approximated by the mean over a finite number M of frequencies, $\omega_j = (\pi j)/(M+1)$ for $j = 0, 1, \ldots, M-1$,

$$\log|\Sigma_\eta^M| = \frac{1}{M}\sum_{j=0}^{M-1} \log 2\pi |g_\xi(\omega_j)|,$$

while the unknown spectral density can be estimated by smoothing the $(k \times k)$ periodogram matrix $I_\xi(\omega; \delta)$ of $\xi_t(\delta)$ over frequencies in the neighborhood of ω,

$$\hat{g}_\xi(\omega; m) = \sum_{|l|<m} W_m(l) I_\xi\Big(\omega + \frac{2\pi l}{T}; \delta\Big).$$

To this end, a bandwidth m and a kernel $W_m(l)$ are applied for which $m \to \infty$, $m/T \to 0$ as $T \to \infty$ and $\sum_{|l|<m} W_m(l) = 1$ hold, and which satisfy also further regularity conditions of Mohanty and Pourahmadi (1996). Taken together, a straightforward estimator for the innovation generalized variance satisfies

$$\log|\widehat{\Sigma}_\eta^M(\delta; m)| = \frac{1}{M}\sum_{j=0}^{M-1} \log\left|2\pi \sum_{|l|<m} W_m(l) I_\xi\Big(\omega_j + \frac{2\pi l}{T}; \delta\Big)\right|. \tag{9}$$

A possible bias correction term for estimating $|\widehat{\Sigma}_\eta^M(\delta; m)|$ is not considered here since it does not change the optimization problem for the resulting estimator,

$$\hat{\delta} = \arg\min_{\delta} \log|\hat{\Sigma}_\eta^M(\delta; m)|. \tag{10}$$

The univariate minimum prediction error variance approach of Proietti and Lütkepohl (2013) results as a special case for $k = 1$ by choosing the uniform kernel $W_m(l) = 1/(2m-1)$ and averaging the smoothed log periodogram over $M = [(T-1)/(2m)]$ frequencies.

In the matrix Box-Cox model, semiparametric estimation of the transformation parameter is more demanding since a scale-preserving normalized transform as in (8) is not available in closed form. In this context, define $\tilde{x}_t := \text{vech}(X_t)$, denote the MBC transformation in vech-space as $\varphi : \tilde{x}_t \mapsto y_t(\delta)$ and the corresponding Jacobi-matrix as

$$\tilde{J}_t(\delta) := \frac{\partial \varphi(\tilde{x}; \delta)}{\partial \tilde{x}'}\bigg|_{\tilde{x}=\tilde{x}_t} \tag{11}$$

for a given observation. A normalized transform is obtained by

$$\tilde{\xi}_t(\delta) := \left|\prod_{s=1}^{T} \tilde{J}_s(\delta)\right|^{-\frac{1}{0.5k(k+1)T}} y_t(\delta).$$

For computational reasons, it is often preferable to work with log determinants by substituting the latter expression into the log of the innovation generalized variance,

$$\log|\widehat{\Sigma}_\eta^M(\delta;m)| = \log|\widehat{\Sigma}_u^M(\delta;m)| - \frac{2}{T}\sum_{t=1}^{T}\log\left|\tilde{J}_t(\delta)\right|, \tag{12}$$

where $|\widehat{\Sigma}_u^M(\delta;m)|$ is the estimated generalized innovation variance of the non-normalized transform $y_t(\delta)$; see (5). When used as an objective for minimization with respect to δ, the Jacobi determinant has to be evaluated numerically.

As a first check if a transformation is relevant for a specific problem at all, it is useful to construct interval estimates for the transformation parameters. If the intervals include unity, then an untransformed approach may be used. Alternatively, matrix logarithmic or square-root models may be a reasonable approximation if the corresponding δ (0 or 0.5, respectively) is contained in the confidence region. Such regions for BC-DC and MBC-RCov transformation parameters can be based on the pivot method, see Casella and Berger (2002, Sec. 9.2.2).

To see how this can be achieved in the current setup, note that for a given δ, the asymptotic distribution of the log innovation generalized variance estimate does not depend on unknown parameters. Mohanty and Pourahmadi (1996, Theorem 3.1(c)) show that under reasonable conditions, for M fixed and $T \to \infty$,

$$\frac{\sqrt{M}}{\sqrt{k\sum_{|j|<m} W_m(j)^2}}\left(\log|\widehat{\Sigma}_\eta^M(\delta;m)| - \log|\Sigma_\eta^M(\delta)|\right) \xrightarrow{d} N(0;1),$$

which is an asymptotically pivotal statistic. A feasible confidence interval for $\log|\Sigma_\eta^M(\delta^*)|$ is given by

$$\log|\widehat{\Sigma}_\eta^M(\hat{\delta};m)| \pm \frac{\sqrt{M}}{\sqrt{k\sum_{|j|<m} W_m(j)^2}} z_{1-\alpha/2}, \tag{13}$$

where $z_{1-\alpha/2}$ is the $1-\alpha/2$ quantile of the standard normal distribution. Following Proietti and Lütkepohl (2013), the confidence region for δ consists of all values δ for which $\log|\widehat{\Sigma}_\eta^M(\delta;m)|$ is contained in (13).

As a practical issue, for both point and interval estimation, the bandwidth parameter m has to be selected. For the univariate approach, a value of $m = 3$ has been found to provide a good balance between bias and variance in Monte Carlo simulations by Proietti and Lütkepohl (2013). In the multivariate case, $m > k$ is required to have positive definite spectral density estimates and hence a nonzero determinant of $\hat{g}(\omega; \delta)$. We try different choices of m in the empirical application below to assess the robustness with respect to the bandwidth choice. Furthermore, we follow Mohanty and Pourahmadi (1996) and set M to the integer part of $0.5T \sum_{|j|<m} W_m(j)^2$, while a uniform kernel is used throughout.

Using this procedure for the realized covariance models introduced in section 2, the transformation parameters can be estimated in a first step. While the MBC-RCov model calls for the multivariate approach, for the BC-DC model either the individual asset variances $X_{ii,t}$ may be used to determine δ_i, $i = 1, \ldots, k$ in turn, or the minimization (10) is carried out for the full vector of realized variances. Leaving dynamic model specification and estimation to a second step makes the analysis computationally convenient: Estimates of the dynamic parameters θ and innovation covariance matrix Σ_u are determined from $y_t(\hat{\delta})$ or $z_t(\hat{\delta})$, respectively. Depending on the dynamic specification, e.g., least squares or Gaussian quasi maximum likelihood methods may be considered.

4 Forecasting and bias correction

4.1 Realized covariance forecasting

Once the parameters of the MBC-RCov model have been appropriately determined, it can be used for forecasting

$$y_{T+h|T}(\delta) := E[y_{T+h}(\delta)|\mathcal{I}_T],$$

where $\mathcal{I}_T$ consists of both returns and realized covariances up to period T. To obtain forecasts of the realized covariance matrices it is necessary to re-transform these predictions into positive definite matrices. Reconstructing a symmetric $(k \times k)$ matrix $Y_{T+h|T}(\delta) = \text{vech}^{-1}(y_{T+h|T}(\delta))$ and applying the inverse of the MBC transform

$$\widetilde{X}_{T+h|T} = H^{-1}(Y_{T+h|T}; \hat{\delta}) \tag{14}$$

may be used as a naïve point forecasts of the realized covariance matrix X_{T+h}.

Due to the nonlinearity of the MBC-transform and its inverse, point forecasts obtained in this way may be severely biased for the conditional mean of X_{T+h}. We therefore propose a simple simulation-based bias correction. Given estimated or pre-specified parameters δ, Σ_u, θ and assuming normally distributed disturbances, we simulate realizations of $y^{(i)}_{T+h}(\delta)$ given $\mathcal{I}_T$ from the model (5) using simulated errors

$$u^{(i)}_t \sim N(0; \Sigma_u), \quad t = T+1, \ldots, T+h, \quad i = 1, \ldots, R.$$

The reverse MBC-transform yields positive definite $X^{(i)}_{T+h}$, $i = 1, \ldots, R$. Averaging over these simulated covariance matrices provides an approximately unbiased point forecast

$$\widehat{X}_{T+h|T} = \frac{1}{R} \sum_{i=1}^{R} X^{(i)}_{T+h}, \tag{15}$$

provided that the normality assumption gives a good description of the actual data generating process. A re-sampling of the model residuals to draw paths of $y^{(i)}_{T+j}(\delta)$ may lead to a procedure which is more robust to deviations from normality. The same procedure can be straightforwardly applied to the BC-DC model and other approaches to transformation-based forecasting as well.

As has been pointed out in section 2, the normality assumption for transformed variables cannot be satisfied whenever $\delta \neq 0$ due to their bounded support. Correspondingly, the re-transformed values of simulated trajectories may not always exist. We circumvent this shortcoming by using draws from a truncated distribution as follows: We first draw paths $u^{(i)}_t$, $t = T, \ldots, T+h$ for $i = 1, \ldots R$ as described above. Whenever a simulated value $y^{(i)}_{T+j}(\delta)$ cannot be re-transformed, we discard the whole trajectory and average over the remaining ones in our bias-correction.

4.2 Forecasting the return distribution

In addition to point forecasts of realized covariance matrices, we consider density prediction of daily returns r_{T+h} conditional on information in period T, denoted $f_r(r_{T+h}|\mathcal{I}_T)$, as this is a key input, e.g., to portfolio decisions and value-at-risk assessment. Joint models of realized covariance and return dynamics have been found beneficial to obtain suitable density forecasts, see Noureldin et al. (2012) or Jin and Maheu (2013). In a Bayesian framework of conditional Wishart models the latter propose computation of such predictive densities that also involves the parameter uncertainty. Our frequentist setup naturally differs from their approach by treating the parameters as fixed and ignoring the estimation error in the computation of density forecasts.

Depending on the intra-day dynamics of returns, the method for computing X_t and the time-span from which daily returns are computed (open-to-close versus close-to-close returns), the unconditional mean of X_t may differ from the unconditional covariance matrix of daily returns, and a re-scaling of the realized measure will be needed. We follow Jin and Maheu (2013) and assume for the daily returns

$$r_t|X_t \sim N(0; X_t^{\frac{1}{2}} \Lambda X_t^{\frac{1}{2}}), \tag{16}$$

where the parameters of the symmetric $(k \times k)$ scaling matrix Λ are estimated by maximum likelihood using daily returns r_t, $t = 1, \ldots, T$. Suppressing the conditioning on $\mathcal{I}_T$ for notational convenience, we obtain draws from $f_X(X_{T+h})$ by simulating $X_{T+h}^{(i)}$ as described in section 4.1 and hence approximate the predictive density at a point r_{T+h} according to

$$f_r(r_{T+h}) = \int f_{r|X}(r_{T+h}|X_{T+h}) f_X(X_{T+h}) \mathrm{d}X_{T+h} \approx \frac{1}{R} \sum_{i=1}^{R} f_{r|X}(r_{T+h}|X_{T+h}^{(i)}), \tag{17}$$

where $f_{r|X}$ is the multivariate normal density of (16).

5 Estimation results

We apply the proposed models to US stock market return and volatility data to assess their usefulness in practice. The data set of Chiriac and Voev (2011) is used which is based on tick-by-tick bid and ask quotes from the NYSE Trade and Quotations (TAQ) database for the period from 2000-01-01 to 2008-07-30 ($T = 2156$). Six liquid stocks are considered, namely (1) American Express Inc., (2) Citigroup, (3) General Electric, (4) Home Depot Inc., (5) International Business Machines and (6) JPMorgan Chase & Co. Chiriac and Voev (2011) describe the computation of the realized covariance matrices from intraday data. They are available from `http://qed.econ.queensu.ca/jae/2011-v26.6/chiriac-voev`.

Estimates of the transformation parameters for the MBC-RCov model are shown in the upper panel of Table 1. They are computed by minimizing (12) for different bandwidths m, using the uniform kernel and setting M to the integer part of $0.5T \sum_{|j|<m} W_m(j)^2$ in (9). Both the unrestricted estimates and the restricted ones under $\delta_1 = \ldots = \delta_k$ are presented. The table reveals that the estimated δ_j are negative but close to zero. The values are all in the range from -0.05 to 0 and are insensitive with regard to the choice of the bandwidth parameters. The confidence intervals which are shown below the respective unrestricted point estimates and

at the right of the restricted ones provide statistical evidence against untransformed, matrix square-root or inverse models. In contrast, the matrix logarithmic model is supported, since all confidence intervals include zero.

To assess the robustness with respect to the estimation method, we also provide the maximum likelihood results for a simple baseline dynamic specification. Details on the estimation approach are given in Appendix A. We assume that the vector of MBC-transformed series follows a VARMA process without dynamic spill-overs, so that each series is represented by an ARMA(p,q)

$$(1-\alpha_{i1}L-\ldots-\alpha_{ip}L^p)(y_{it}(\delta)-\mu_i) = (1+\phi_{i1}L+\ldots+\phi_{iq}L^q)u_{it}(\delta),\ i=1,\ldots,k(k+1)/2. \quad (18)$$

If ARMA models are estimated for $y_{it}(\hat{\delta})$, $i=1,\ldots,k$, with δ determined by the semiparametric method, the BIC favors orders (p,q) of (1,1), (2,1) or, less frequently, (2,2) for all but 2 of the 21 series. We choose the diagonal VARMA(2,1) model as it reconciles these outcomes with a quest for parsimony. With three dynamic parameters per series, this model is similar in complexity to recent successful approaches to daily (co-)variance modeling; see, e.g., Chiriac and Voev (2011). As in their model, Granger-causality relations between the series are excluded. The maximum likelihood estimates, given in the lower panel of Table 1, are close to the semiparametric ones.

Estimation results for the transformation parameters of univariate asset variances, which are utilized in the BC-DC approach, are provided in Table 2. The Box-Cox parameters are again negative with small magnitude. In contrast to the MBC-RCov case, the univariate estimates are spread over the range $[-0.1;0]$; the multivariate estimates are closer to zero again. By including zero, the confidence intervals suggest that a linear time series model may simply be applied to the log realized variances and z-transformed correlations. This conclusion does not change when the maximum likelihood estimator for the diagonal VARMA(2,1) model is considered. Again, this is the specification which is individually favored by the BIC for most series.

Figure 1 shows two time series plots of each, the raw series X_t, the MBC-transformed $y_t(\hat{\delta})$ (with semiparametric estimates $\hat{\delta}_j$ and $m=42$) and the BC-DC transformed $z_t(\hat{\delta})$ (with univariate semiparametric estimates $\hat{\delta}_j$ and $m=3$) in the left two panels. While large positive outliers occur and volatility strongly co-moves with the level of the untransformed variances and covariances, the distribution of the transformed series is more stable and symmetric around their persistent movements. Interestingly, the dynamics of the first diagonal entry of the

MBC series, $y_{1,t}(\hat{\delta})$, are very similar to the univariate transformed variance $z_{1,t}(\hat{\delta})$, while the nondiagonal MBC entry $y_{2,t}(\hat{\delta})$ evolves in close parallel to the corresponding z-transformed correlation $z_{7,t}(\hat{\delta})$. This reflects the finding of Gribisch (2013, p.4) who found and discussed the closeness of the matrix logarithm's diagonal and off-diagonal elements to log variances and correlations, respectively.

We assess the in-sample success of various nonlinear transforms for our dataset. The appropriateness of a transform for modeling purposes can be seen by a parsimonious ARMA representation for the transformed variable. Additionally, the stabilization of conditional variances, i.e. the conditional homoscedasticity of the residuals, is an important goal of the transformation. Further, the approximate normality of transformed variable or model residuals are frequently stated as the motivation for a transformation-based approach. To evaluate these goals, diagnostic residual tests are carried out for our benchmark VARMA(2,1) specification which is applied to different transforms. As straightforward and familiar choices, we use univariate Ljung-Box tests both on raw residuals and on squared residuals to check serial correlation and conditional heteroscedasticity, respectively, while Jarque-Bera tests are used to detect deviations from normality. In our comparison, we consider the raw realized variances and covariances (vech), the nonzero terms of the Cholesky factors (chol) as well as the unique terms of the symmetric matrix square-root (sym-root), of the matrix logarithm (mlog) and of the inverse covariance matrices (inverse).

The results of these diagnostic tests, more precisely the number of rejections across series for different significance levels, is given in Table 3. Transformations of the realized covariance matrices which are not appropriate may be detectable by the failure of linear time series models to produce serially uncorrelated residuals. Therefore, consider the test for residual autocorrelation, given in the upper panel. Both the MBC-RCov and the BC-DC models have a non-negligible fraction of rejections. There remains significant autocorrelation even at the 0.1% level for one of the 21 series. Autocorrelation is modest compared to some of the other transformations, e.g., the Cholesky factorization, however. For the latter, the p-values are smaller than 1% in 9 cases and the majority (14 out of 21) of the residual series exert significant autocorrelation at the 5% level. The matrix logarithm transform is the hardest competitor but still exceeds the BC-DC model with respect to the number of rejections (4 versus 2 at the 1% level). A better fit of the models could be obtained by using larger VARMA model orders globally, or by specifying the ARMA models individually for each series. We have consistently chosen an intuitive and sparse specification to enhance comparability with other model classes in light of the out-of-sample forecasting study carried out below.

Compared to the other transforms and special cases, the BC-DC and even more the MBC-RCov model (2 and 1 rejection, respectively, at the 1% level) succeed in stabilizing the residual variance. Except for the mlog-transform, all other models suffer from extreme conditional heteroscedasticity with very high rejection rates when autocorrelation of the squared residuals is tested.

Normality of the model residuals is frequently rejected for the Box-Cox specifications, namely 10 times out of 21 at the 1% level. It thus performs worse than the matrix logarithm. Given that the matrix logarithm is contained in the family of MBC-transforms it turns out that when choosing the transformation parameters for our dataset, we face a tradeoff between obtaining linear time series models, stabilizing variance and yielding normally distributed variables. The estimated $\delta_j < 0$ thus prioritize the former goals but fail with respect to the latter. Still, the normal approximation is strongly improved compared to the raw variance and covariance series. Kernel density plots of the model residuals in the right panel of Figure 1 show the approximate normality of the transformed series as compared to the untransformed ones.

6 Forecast comparison

We assess the forecast performance of the MBC-RCov and the BC-DC models, also in light of popular competitor methods. To this end, we use the data set introduced in section 5 and conduct a quasi out-of-sample forecast exercise, recursively using a pre-specified window of data for parameter estimation and forecasting, and then, subsequently, evaluating the forecasts against realized data outside that range.

We address the following questions in the forecasting exercise. To take a closer look at the methods introduced in section 4.1, we first assess whether bias-correction leads to a significant improvement of the forecasts. Further, for both the MBC-RCov and the BC-DC model, we evaluate which value of the transformation parameter δ dominates in terms of out-of-sample precision and whether the estimates presented in section 5 are also superior out of sample. We are also interested in whether the dynamic correlation specification outperforms the matrix transform or vice versa.

Another main objective of the study is to assess the value of the transformation-based approach as compared to other methods. As a recently suggested and popular competitor we consider the class of models that assume conditionally Wishart distributed realized covariance matrices. More specifically, two models are included in our baseline comparison, the Conditional Autoregressive Wishart (CAW) model of Golosnoy et al. (2012) and the Conditional Autoregressive Wishart Dynamic Conditional Correlation (CAW-DCC) specification of Bauwens et al. (2012).

6.1 Models and setup

To tackle the questions above, we apply the baseline diagonal VARMA(2,1) specification for the MBC-RCov and BC-DC model and apply a grid of fixed values for the transformation parameter which seem relevant for a specific comparison, e.g., $\delta_1 = \ldots = \delta_k \in \{-0.1, -0.05, 0, 0.5, 1\}$. The other model parameters are re-estimated for each estimation window.

For the Wishart models used as benchmarks, the distributional assumption is

$$X_t|\mathcal{I}_{t-1} \sim W_n(\nu, S_t/\nu), \tag{19}$$

where W_n denotes the central Wishart density, ν is the scalar degrees of freedom parameter and S_t/ν is a $(k \times k)$ positive definite scale matrix, which is related to the conditional mean of X_t by $E[X_t|\mathcal{I}_{t-1}] = S_t$. The baseline CAW(p,q) model of Golosnoy et al. (2012) specifies the conditional mean as

$$S_t = CC' + \sum_{j=1}^{p} B_j S_{t-j} B_j' + \sum_{j=1}^{q} A_j X_{t-j} A_j', \tag{20}$$

C, B_j and A_j denoting $(k \times k)$ parameter matrices, while the CAW-DCC model of Bauwens et al. (2012) employs a decomposition

$$S_t = H_t P_t H_t', \tag{21}$$

where H_t is diagonal and P_t is a well-defined correlation matrix. As a sparse and simple DCC benchmark we apply univariate realized GARCH(p_v,q_v) specifications for the realized variances

$$H_{ii,t}^2 = c_i + \sum_{j=1}^{p_v} b_{i,j}^v H_{ii,t-j}^2 + \sum_{j=1}^{q_v} a_{i,j}^v X_{ii,t-j}, \tag{22}$$

along with the 'scalar Re-DCC' model (Bauwens et al.; 2012) for the realized correlation matrix R_t,

$$P_t = \bar{P} + \sum_{j=1}^{p_c} b_j^c P_{t-j} + \sum_{j=1}^{q_c} a_j^c R_{t-j}. \tag{23}$$

The diagonal CAW(p,q) and the CAW-DCC(p,q) specification with $p = p_v = p_c = 2$ and $q = q_v = q_c = 1$ are selected since they are similar in complexity to the diagonal VARMA(2,1) model and provide a reasonable in-sample fit among various order choices.

For a given forecasting method, the evaluation is carried out as follows: We split the available data in a sample $X_1, \ldots, X_{1508}$ which is used only for estimation and an evaluation sample $X_{1509}, \ldots, X_{2156}$. For each $T' \in [1508; 2156-h]$, the model is estimated using a rolling sample $X_{T'-1507}, \ldots, X_{T'}$ of 1508 observations and forecasts of $X_{T'+h}$, $h = 1, 5, 10, 20$, are computed. For the transformation-based forecasts, we consider both the naïve forecast $\widetilde{X}_{T'+h|T'}$, based on (14), and the bias-corrected forecast $\widehat{X}_{T'+h|T'}$, see (15), using $R = 1000$ simulated realizations. In addition, density forecasts of the returns $r_{T'+h}$ given $\mathcal{I}_{T'}$ are computed using the same simulated trajectories as for the bias-corrected covariance forecasts.

As outlined in section 4.1, in the simulations we discard trajectories where the transformation from $y_t^{(i)}$ to $X_t^{(i)}$ is not well-defined. Additionally, to attenuate the effect of extreme outliers in the simulated paths, we replace an element of the bias-corrected covariance matrix forecast by the uncorrected forecast if the fraction between the two exceeds 5 in absolute value. Such a procedure reflects a practically feasible plausibility check. Both modifications of the forecasts are needed only for small values of the transformation parameters $\delta \leq -0.25$ in our study which are anyway inconsistent with the empirical interval estimates presented above.

We compare the forecasting models by presenting average losses, i.e. risks, over the evaluation period. To gain insights about statistical significance of the differences, model confidence sets (MCS) are constructed following Hansen et al. (2011) using the MulCom package (Hansen and Lunde; 2010) in Ox Console Version 6.21. The Max-t statistic is bootstrapped with a block lengths of $d = \max\{5, h\}$ and 10000 iterations. A confidence level of 90% is used throughout.

6.2 Baseline results

We begin with an evaluation of the forecasting performance using a simple *squared prediction error* loss function, evaluated using the true realized covariance matrices. For $h = 1, 5, 10, 20$ and $T' = 1508, \ldots, 2156-h$, we compute the period loss as the Frobenius norm of the forecast error

$$LF^{\{s\}}_{T',h} = \sum_{i=1}^{k}\sum_{j=1}^{k}\left(X_{ij,T'+h} - X^{\{s\}}_{ij,T'+h|T'}\right)^2, \tag{24}$$

where $X^{\{s\}}_{T'+h|T'}$ is a covariance matrix forecast obtained from one of the different methods.

To assess the value of the bias-correction we compare the risk

$$\overline{LF}^{\{s\}}_{h} = \frac{1}{2156-h-1507}\sum_{T'=1508}^{2156-h} LF^{\{s\}}_{T',h}$$

of the corrected forecasts to the naïve ones by calculating the fraction of the two for the different models and transformation parameters. The results are given in Table 4. Despite the adjustment of miss-behaved bias-corrected forecasts outlined in section 6.1, the simulation-based forecasts are worse than the uncorrected ones for $\delta \leq -0.25$. This is most pronounced for the MBC-RCov model and for short horizons. In such cases, the normality assumption provides a poor description of the transformed variables and hence the simulated $y_t^{(i)}$ series do not produce well-behaved re-transformed forecasts.

The valuation of bias correction changes fundamentally when $\delta \geq -0.1$ is considered. This is the empirically relevant span as the estimates of section 5 suggest. The simulation-based procedure leads to marked improvements of the forecasts. The reduction in risks for the MBC-RCov model is as high as 12% for $h = 1$ and $\delta = -0.1$; it gradually reduces as δ approaches one. There, the MBC-transform corresponds to the raw covariance series and bias-correction is not needed. This broad picture is reflected also by the BC-DC transformed series, where $\delta = 0.5$, corresponding to a model of realized standard deviations, is minimally prone to bias. For $\delta = 1$, when untransformed realized variances are approximated by a Gaussian process in the simulations, the latter fail to reduce bias, and even devastate the forecast accuracy.

Having shown the usefulness of bias correction, we now turn to a comparison of bias-corrected forecasts of the proposed models along with the CAW and CAW-DCC specification for which a correction is not needed. The corresponding Frobenius risks are shown in the left part of Table 5. The boldface numbers, which indicate the best-performing model for each horizon, show that the BC-DC specification with $\delta \in \{-0.1, -0.05\}$ emerges as favorable for all horizons.

The asterisks indicate models contained in the 90% model confidence set for a given horizon h. The MBC-RCov forecasts are contained only for a few specific values of δ. For one-step forecasts, only the matrix square root transformed ($\delta = 0.5$) and untransformed ($\delta = 1$)

forecasts cannot be rejected. For some other horizons, no MBC-RCov specification ($h = 5$) or only those with negative δ ($h = 20$) resist a rejection. Neither the matrix logarithm ($\delta = 0$) nor the semiparametric estimates of δ provide a reasonable performance which is robust with respect to the chosen horizon. The BC-DC model forecasts are elements of the MCS for a wide range of transformation parameters including the estimates from section 5 as well as $\delta = 0$. This holds for all considered forecast horizons. A comparison to the Wishart models reveals that despite their larger risk, both models cannot be rejected by the MCS approach except for the two-weeks horizon $h = 10$.

These conclusions about superiority change when the density forecasts are considered. We evaluate the forecasts by the logarithmic scoring rule; see, e.g., Gneiting et al. (2008),

$$LD_{T',h}^{\{s\}} = -\log f_r^{\{s\}}(r_{T'+h}|\mathcal{I}_{T'}), \tag{25}$$

computing negative logarithms of the density forecast $f_r^{\{s\}}$, evaluated at the realized daily returns h periods ahead. The logarithmic rule is a strictly proper scoring rule, rewarding careful and honest assessments (Gneiting and Raftery; 2007). It is local in the sense that no point of $f_r^{\{s\}}$ other than the realized return is evaluated, which is also an intuitively and computationally appealing property.

The results are given in the right part of Table 5. Here, the MBC-RCov with $\delta \approx 0$ outperforms the BC-DC model; significantly for $h = 1$ as the MCS consists of only one specification there. For larger horizons, the differences are less pronounced and both models with transformation parameters close to zero seem appropriate.

To understand this outcome, note that the whole density of $X_{T'+h}$ given $X_{T'}$ is involved in computing the conditional return density forecasts. The matrix logarithmic model stands out from its competitors in yielding relatively close-to-Gaussian residuals, as has been seen in section 3. Since we use the truncated normal distribution in our simulation of $X_{T'+h}^{(i)}$, the different model rankings with respect to covariance forecasts and return density forecasts can be understood in light of this finding.

Also the Wishart models are rejected using the log density metric. We regard this as evidence that the Gaussianity assumption for transformed realized covariances provides a better approximation than the Wishart specification — at least when it comes to forecasting the return distribution.

To conclude the baseline results, the BC-DC model with small negative δ outperforms in terms of covariance matrix forecasting, while the MBC-RCov model with $\delta \approx 0$ emerges when the aim is forecasting the return density. The Wishart models are outperformed significantly in the latter case.

6.3 Robustness regarding model specification

Up to now, the results for MBC-RCov and BC-DC forecasts are based on a simple diagonal VARMA(2,1) specification. We check whether our conclusions with regard to the data transformations remain intact for other models which have been used for volatility dynamics.

We first assess whether the choice of VARMA order matters for our conclusion. To this end, we compute forecasts and risks also for other specifications. The result for the Frobenius loss and $h = 10$ is exemplarily shown in Figure 2 and mirrors the result for other horizons. To make the figures comparable, here and henceforth, the risks are plotted as a fraction of a common benchmark, the BC-DC model with VARMA(2,1) dynamics and $\delta = 0$. It turns out that the VARMA(2,1) specification is among the best choices for most of the different transformation parameters. Importantly, the conclusion about favorable transforms does not interfere with the choice of model orders.

Daily financial volatility is often associated with a long memory behaviour, so we also include such models to our robustness checks. As a first alternative, we consider the heterogeneous autoregressive model of Corsi (2009). Lags of $y_{it}(\delta)$, averaged over 1, 5 and 20 trading days in the process

$$y_{it}(\delta) = c_i + \alpha_{i1} y_{i,t-1}(\delta) + \alpha_{i5}\Big(\frac{1}{5}\sum_{j=1}^{5} y_{i,t-j}(\delta)\Big) + \alpha_{i20}\Big(\frac{1}{20}\sum_{j=1}^{20} y_{i,t-j}(\delta)\Big) + u_{it} \tag{26}$$

introduce long-memory-like persistence. The parameters are estimated by least squares. In contrast, Chiriac and Voev (2011) use a flexible fractionally integrated vector ARMA (VARFIMA) specification with “real” long memory behavior. We also follow their approach but do not restrict the memory parameter to be the same across series, and hence estimate series-specific parameters $\theta_i = (d_i, \alpha_{i1}, \ldots, \alpha_{ip}, \phi_{i1}, \ldots, \phi_{iq}, \mu_i)$ of

$$(1 - \alpha_{i1}L - \ldots - \alpha_{ip}L^p)(1 - L)^{d_i}(y_{it}(\delta) - \mu_i) = (1 + \phi_{i1}L + \ldots + \phi_{iq}L^q)u_{it}. \tag{27}$$

Again, correlation between the series is introduced only through the noise covariance matrix

Σ_u. Like Chiriac and Voev (2011) we set $p = q = 1$ which gives the same model complexity as in our benchmark VARMA(2,1).

Overall, the VARFIMA setup provides smaller forecast errors than the VARMA benchmark, while the HAR is outperformed by both competitors; see the results in Figure 3. The excellent results for the ARFIMA model are in line with the results of Chiriac and Voev (2011). Further gains may be attainable by considering more sophisticated models, e.g., taking possible dynamic spillovers, factor structures and structural breaks into account. While a comprehensive comparison of different dynamic specifications is beyond the scope of this paper, we direct attention to the relative benefits of the various transformations for a given model. The relative rankings remain remarkably unchanged, independently of the dynamic specification. Again, the BC-DC model with small negative δ_j stands out.

Lastly, we compare our transformation-based approach to other models of the conditional Wishart family. To this end, we conduct a comparison of several diagonal CAW(p,q) models and CAW-DCC(p,q) models with different orders p and q. Additionally, the component models proposed by Jin and Maheu (2013) are considered. Regarding the latter, we estimate a Wishart Additive Component (CAW-ACOMP) model. The distributional assumption (19) is complemented by

$$S_t = CC' + \sum_{j=1}^{K} B_j \odot \Gamma_{t,l_j}, \qquad B_j = b_j b_j', \qquad \Gamma_{t,l} = \frac{1}{l} \sum_{i=0}^{l-1} X_{t-i}, \tag{28}$$

where $\odot$ denotes the elementwise (Hadamard) product and, analogously to the HAR model, past averages of the covariances enter the conditional mean equation in a linear manner. Similarly, such lower frequency components are also involved in the Wishart Multiplicative Component (CAW-MCOMP) model

$$S_t = \left[\prod_{j=K}^{1} \Gamma_{t,l_j}^{\frac{\gamma_j}{2}} \right] CC' \left[\prod_{j=1}^{K} \Gamma_{t,l_j}^{\frac{\gamma_j}{2}} \right], \tag{29}$$

which we also assess in our study. As for the HAR model, we set $K = 3$ and average over $l_1 = 1$, $l_2 = 5$ and $l_3 = 20$ past observations. In accordance with all other Wishart models considered so far, the parameters are estimated by maximum likelihood for all rolling samples.

The results which are shown in Figure 4 for the Frobenius loss are clear-cut. None of the models outperforms the BC-DC benchmark, irrespective of the forecasting horizon. This is indicated by the relative risks that are above one for all models. In contrast, the ranking among the Wishart models varies with the horizon. At least for the smaller horizons, the chosen benchmark orders $p = 2$ and $q = 1$ correspond to well-performing models. The component models show an ambiguous figure. The additive model does well for most horizons, but the multiplicative approach is worthwhile only for rather long-term forecasts ($h = 20$).

Overall, the robustness checks find that the results of the baseline setup remain qualitatively unchanged also when other dynamic models, both for the Box-Cox models and for the Wishart family, are taken into account. Among the considered alternatives, long memory BC-DC models are the most relevant direction for improvements.

6.4 Robustness regarding loss function

So far we have focussed on the Frobenius norm of the forecast error when evaluating the covariance matrix forecasts. In the matrix case, there are several other loss functions which may be appropriate for different practical forecasting situations; see, e.g., Laurent et al. (2013) for a discussion. In our out-of-sample study, we additionally consider the Stein distance

$$LS_{T',h} = \operatorname{tr}\left[X^{-1}_{T'+h|T'} X_{T'+h}\right] - \log\left|X^{-1}_{T'+h|T'} X_{T'+h}\right| - k, \tag{30}$$

and the asymmetric loss

$$L3_{T',h} = \frac{1}{6}\operatorname{tr}\left[X^{3}_{T'+h|T'} - X^{3}_{T'+h}\right] - \frac{1}{2}\operatorname{tr}\left[X^{2}_{T'+h|T'}(X_{T'+h} - X_{T'+h|T'})\right], \tag{31}$$

which is used by Laurent et al. (2011). Forecast comparisons based on LF, LS and $L3$ may differ because LS penalizes underpredictions more heavily than LF while overpredictions are more influential with the $L3$ loss, see Laurent et al. (2011), section 2.3. Additionally, the loss functions differ in their relative importance of high versus low volatility periods since only the Stein distance is homogeneous of order 0 and hence scale invariant.

The results of the evaluation with the Frobenius norm is replicated using both the LS and the $L3$ norms. Again, as Table 6 reveals, the BC-DC models perform better than their MBC-RCov counterparts and their forecasting superiority is not rejected with zero or small negative transformation parameter. The CAW models are statistically rejected in some cases, even if the power of the MCS procedure appears small for the L3 loss.

A reasonable loss function may also be chosen to involve the economic cost of prediction errors. A risk-averse investor may be interested in the variance of an ex-ante minimum-variance portfolio (MVP) which is computed from the covariance matrix forecast. Using the realized variance of the MVP as the ex-post loss, we therefore consider

$$LMV_{T',h} = w' X_{T'+h} w, \qquad \text{where} \qquad w = (\iota' X_{T'+h|T'} \iota)^{-1} X_{T'+h|T'} \iota, \tag{32}$$

where $\iota = (1, \dots, 1)'$. Alternatively, the squared daily return $w' r_{T'+h} r'_{T'+h} w$ of the ex-ante MVP is used instead of the realized variance.

Table 7 shows rather inconclusive results. The discriminating power is weak for these two losses, so that many models are included in the model confidence sets. Notably, however, BC-DC with $\delta = 0$ outperforms in three out of eight horizon-loss combinations and is always included in the MCS.

To summarize, the forecasting results are unchanged if other dynamic models are considered and reveal relatively little ambiguity also with alternative loss functions. Overall, the Box-Cox dynamic correlation specification with log variances ($\delta = 0$) seems to be a good and robust choice in practice. It is close to the best performing model for most horizons and with regard to many of the evaluation criteria. The MBC-RCov model, however, has a superior forecasting performance for specific criteria such as predictive densities. Further research appears fruitful to further clarify these facts, e.g., in light of datasets for different asset classes.

7 Conclusion

We have proposed two new approaches to multivariate realized volatility modeling and applied them to US stock market data. The empirical results, including an out-of-sample forecasting comparison, seem promising, also in comparison to the main competitors, the conditional autoregressive Wishart model of Golosnoy et al. (2012) and several variants thereof. Our assessment of various transformation parameters supports a convenient special case of our Box-Cox approaches: the use of standard linear time series models to a multivariate time series of log realized variances and z-transformed correlations. Its appropriateness can be easily checked for a specific dataset using the inferential methods introduced in this paper.

The present study leaves significant questions for further research. With a focus on forecasting, investigating more advanced dynamic specifications appears worthwhile, possibly including dynamic spillovers and structural changes along with the long memory dynamics briefly considered in this paper. Additionally, in applications to data sets of higher dimensions, our approach allows the assessment of cross-sectional properties such as factor structures in a methodologically and computationally straightforward setup. In addition to the realized volatility setup with utilization of intraday data, our models are also relevant for the study of multivariate stochastic volatility based on a latent covariance matrix specification.

References

Amemiya, T. and Powell, J. L. (1981). A comparison of the Box-Cox maximum likelihood estimator and the non-linear two-stage least squares estimator, *Journal of Econometrics* **17**(3): 351–381.

Andersen, T. G., Bollerslev, T., Diebold, Francis, X. and Labys, P. (2003). Modelling and forecasting realized volatility, *Econometrica* **71**(2): 579–625.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariation: High frequency based covariance, regression, and correlation in financial economics, *Econometrica* **72**(3): 885–925.

Bauer, G. H. and Vorkink, K. (2011). Forecasting multivariate realized stock market volatility, *Journal of Econometrics* **160**(1): 93–101.

Bauwens, L., Storti, G. and Violante, F. (2012). Dynamic conditional correlation models for realized covariance matrices, *CORE Discussion Paper 2012/60.*

Box, G. E. P. and Cox, D. R. (1964). An analysis of transformations, *Journal of the Royal Statistical Society. Series B* **26**(2): 211–252.

Box, G. E. P. and Jenkins, G. M. (1970). *Time Series Analysis: Forecasting and Control*, Holden Day.

Casella, G. and Berger, R. L. (2002). *Statistical Inference*, Thomson Learning.

Chiriac, R. and Voev, V. (2011). Modelling and forecasting multivariate realized volatility, *Journal of Applied Econometrics* **26**(6): 922–947.

Corsi, F. (2009). A simple approximate long memory model of realized volatility, *Journal of Financial Econometrics* **7**: 174–195.

Draper, N. R. and Cox, D. R. (1969). On distributions and their transformation to normality, *Journal of the Royal Statistical Society. Series B (Methodological)* **31**(3): 472–476.

Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models, *Journal of Business and Economic Statistics* **20**(3): 339–350.

Gneiting, T. and Raftery, A. E. (2007). Strictly proper scoring rules, prediction, and estimation, *Journal of the American Statistical Association* **102**: 359–378.

Gneiting, T., Stanberry, L. I., Grimit, E. P., Held, L. and Johnson, N. A. (2008). Assessing probabilistic forecasts of multivariate quantities, with an application to ensemble predictions of surface winds, *TEST* **17**(2): 211–235.

Golosnoy, V., Gribisch, B. and Liesenfeld, R. (2012). The conditional autoregressive Wishart model for multivariate stock market volatility, *Journal of Econometrics* **167**(1): 211–223.

Golosnoy, V. and Herwartz, H. (2012). Dynamic modeling of high-dimensional correlation matrices in finance, *International Journal of Theoretical and Applied Finance* **15**(5): 1–22.

Goncalves, S. and Meddahi, N. (2011). Box-Cox transforms for realized volatility, *Journal of Econometrics* **160**(1): 129–144.

Gribisch, B. (2013). A latent dynamic factor approach to forecasting multivariate stock market volatility, *Beiträge zur Jahrestagung des Vereins für Socialpolitik 2013: Wettbewerbspolitik und Regulierung in einer globalen Wirtschaftsordnung - Session: Volatility, F01-V2*.

Halbleib, R. and Voev, V. (2011). Forecasting covariance matrices: A mixed frequency approach, *Working Papers ECARES 2010-002*, Universite Libre de Bruxelles.

Han, A. K. (1987). A non-parametric analysis of transformations, *Journal of Econometrics* **35**: 191–209.

Hansen, P. R. and Lunde, A. (2010). *MulCom 2.00 - Econometric Toolkit for Multiple Comparisons*.

Hansen, P. R., Lunde, A. and Nason, J. (2011). The model confidence set, *Econometrica* **79**(2): 453–497.

Härdle, W. and Simar, L. (2007). *Applied Multivariate Statistical Analysis*, Springer.

Hautsch, N., Kyj, L. M. and Malec, P. (2014). Do high-frequency data improve high-dimensional portfolio allocations? *Journal of Applied Econometrics*, forthcoming.

Higgins, M. L. and Bera, A. (1992). A class of nonlinear ARCH models, *International Economic Review* **33**: 137–158.

Jin, X. and Maheu, J. M. (2013). Modelling realized covariances and returns, *Journal of Financial Econometrics* **11**(2): 335–369.

Jones, R. H. (1976). Estimation of the innovation generalized variance of a multivariate stationary time series, *Journal of the American Statistical Association* **71**(354): 386–388.

Laurent, S., Rombouts, J. V. K. and Violante, F. (2011). On the forecasting accuracy of multivariate GARCH models, *Journal of Applied Econometrics* **27**(6): 934–955.

Laurent, S., Rombouts, J. V. K. and Violante, F. (2013). On loss functions and ranking forecasting performances of multivariate volatility models, *Journal of Econometrics* **173**(1): 1–10.

Mohanty, R. and Pourahmadi, M. (1996). Estimation of the generalized prediction error variance of a multiple time series, *Journal of the American Statistical Association* **91**(433): 294–299.

Noureldin, D., Shephard, N. and Sheppard, K. (2012). Multivariate high-frequency-based volatility (HEAVY) models, *Journal of Applied Econometrics* **27**(6): 907–933.

Priestley, M. B. (1982). *Spectral Analysis and Time Series.*

Proietti, T. and Lütkepohl, H. (2013). Does the Box–Cox transformation help in forecasting macroeconomic time series?, *International Journal of Forecasting* **29**(1): 88–99.

Velilla, S. (1993). A note on the multivariate Box-Cox transformation to normality, *Statistics and Probability Letters* **17**: 259–263.

Yu, J., Yang, Z. and Zhang, X. (2006). A class of nonlinear stochastic volatility models and its implications for pricing currency options, *Computational Statistics and Data Analysis* **51**: 2218–2231.

Zhang, X. and King, M. L. (2008). Box-Cox stochastic volatility models with heavy-tails and correlated errors, *Journal of Empirical Finance* **15**(3): 549–566.

A Maximum likelihood estimation

This appendix describes Maximum Likelihood estimation of the MBC-RCov model as conducted in section 5 and shown in the lower panel of Table 1. Although the transformed series cannot be exactly Gaussian due to the bounded support, we use

$$u_t \sim NID(0; \Sigma_u)$$

as an approximating auxiliary assumption for parameter estimation, alongside a dynamic model specification (5), the VARMA model (18) in our case. Under this assumption, the conditional distribution of $y_t(\delta)$ is also Gaussian $N(\mu_t; \Sigma_u)$ with conditional mean $\mu_t(\theta)$ determined by the time series model. Denoting, as in section 3, the vector of untransformed variances and covariances by $\tilde{x}_t := \text{vech}(X_t)$, the MBC transformation in vech-space as $y_t(\delta) = \varphi(\tilde{x}_t; \delta)$ and the Jacobi matrix as $\tilde{J}_t(\delta)$ (see (11)), the joint density of $\tilde{x}_t$ is given by

$$f_x(\tilde{x}_t; \delta, \Sigma_u, \theta) = |\tilde{J}_t(\delta)||2\pi\Sigma_u|^{-\frac{1}{2}} \exp\left\{ -\frac{1}{2}(\varphi(\tilde{x}_t; \delta) - \mu_t(\theta))'\Sigma_u(\varphi(\tilde{x}_t; \delta) - \mu_t(\theta)) \right\},$$

see, e.g., Härdle and Simar (2007, section 3.7).

The log-likelihood, conditional on pre-sample values $x_0, x_{-1}, \ldots$, is hence

$$l(\delta, \Sigma_u, \theta) = \sum_{t=1}^{T} \log |\tilde{J}_t(\delta)| - \frac{T}{2} \log |\Sigma_u| - \frac{1}{2} \sum_{t=1}^{T} (\varphi(\tilde{x}_t; \delta) - \mu_t(\theta))'\Sigma_u^{-1}(\varphi(\tilde{x}_t; \delta) - \mu_t(\theta)). \quad (33)$$

Given δ and θ, the unrestricted maximum likelihood estimator of Σ_u is computed by $\Sigma_u(\theta, \delta) = \frac{1}{T}\sum_{t=1}^{T}(\varphi(\tilde{x}_t; \delta) - \mu_t(\theta))(\varphi(\tilde{x}_t; \delta) - \mu_t(\theta))'$ which can be plugged into (33) to obtain the concentrated likelihood

$$l^c(\delta, \theta) = \sum_{t=1}^{T} \log |\tilde{J}_t(\delta)| - \frac{T}{2} \log \left| \frac{1}{T} \sum_{t=1}^{T} (\varphi(\tilde{x}_t; \delta) - \mu_t(\theta))(\varphi(\tilde{x}_t; \delta) - \mu_t(\theta))' \right|. \quad (34)$$

In practice, a further concentration step seems worthwhile. Compute, for a given parameter vector δ, the maximum likelihood estimator for the dynamic parameters θ. The term $\log |\tilde{J}_t(\delta)|$ does not affect this optimization so that computation of the Jacobian can be suppressed. The concentrated likelihood (34) is then maximized with respect to δ only, with the Jacobian numerically computed at each likelihood evaluation.

Tables

Table 1: Estimates of the transformation parameters $\delta_1, \ldots, \delta_k$ for the covariance eigenvalues in the MBC-RCov model based on (3) and (5). The semiparametric estimator outlined in section 3 as well as the Maximum Likelihood estimator for the VARMA(2,1) specification (18) (see Appendix A) are applied to the realized covariance matrices. Estimates under the constraint $\delta_1 = \ldots = \delta_k$ ("restricted") and without this constraint ("unrestricted") are presented.

m	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6
			Semiparametric unrestricted			
21	-0.0306	-0.0397	-0.0180	-0.0144	-0.0176	-0.0266
	[-0.11; 0.06]	[-0.11; 0.03]	[-0.09; 0.05]	[-0.09; 0.06]	[-0.09; 0.06]	[-0.12; 0.07]
42	-0.0265	-0.0315	-0.0272	-0.0271	-0.0143	-0.0231
	[-0.11; 0.06]	[-0.10; 0.03]	[-0.09; 0.04]	[-0.09; 0.06]	[-0.10; 0.06]	[-0.12; 0.07]
63	-0.0230	-0.0274	-0.0206	-0.0157	-0.0162	-0.0253
	[-0.10; 0.06]	[-0.11; 0.03]	[-0.08; 0.06]	[-0.09; 0.06]	[-0.09; 0.06]	[-0.12; 0.07]
			Semiparametric restricted $\delta_1 = \ldots = \delta_k$			
21			-0.0328	[-0.12; 0.05]		
42			-0.0270	[-0.12; 0.05]		
63			-0.0252	[-0.12; 0.05]		
			Maximum Likelihood			
unrestr.	-0.0238	-0.0279	-0.0225	-0.0180	-0.0213	-0.0316
restr.			-0.0239			

Table 2: Estimates of the transformation parameter $\delta_1, \ldots, \delta_k$ for the realized variances in the BC-DC model based on (7) and (5). The semiparametric estimator outlined in section 3 as well as the Maximum Likelihood estimator for the VARMA(2,1) specification (18) (see Appendix A). The variances are ordered as (1) American Express Inc., (2) Citigroup, (3) General Electric, (4) Home Depot Inc., (5) International Business Machines and (6) JPMorgan Chase & Co. Estimates under the constraint $\delta_1 = \ldots = \delta_k$ ("restricted") and without this constraint ("unrestricted") are presented, while this restriction is not possible for the univariate estimators.

m	δ_1	δ_2	δ_3	δ_4	δ_5	δ_6
			Semiparametric univariate			
1	-0.0616	-0.0934	-0.0715	-0.0346	-0.0762	-0.0569
	[-0.20; 0.07]	[-0.23; 0.05]	[-0.23; 0.08]	[-0.22; 0.14]	[-0.25; 0.09]	[-0.19; 0.08]
3	-0.0619	-0.0888	-0.0645	-0.0382	-0.0758	-0.0525
	[-0.20; 0.07]	[-0.22; 0.04]	[-0.22; 0.09]	[-0.22; 0.14]	[-0.24; 0.08]	[-0.19; 0.08]
10	-0.0613	-0.0832	-0.0592	-0.0353	-0.0708	-0.0533
	[-0.19; 0.07]	[-0.22; 0.05]	[-0.21; 0.09]	[-0.22; 0.13]	[-0.24; 0.09]	[-0.19; 0.07]
			Semiparametric multivariate unrestricted			
6	-0.0093	-0.0354	-0.0433	-0.0165	-0.0501	-0.0127
	[-0.2; 0.17]	[-0.2; 0.12]	[-0.25; 0.15]	[-0.29; 0.24]	[-0.3; 0.18]	[-0.18; 0.14]
12	-0.0189	-0.0365	-0.0346	-0.0114	-0.0435	-0.0095
	[-0.2; 0.16]	[-0.2; 0.12]	[-0.24; 0.16]	[-0.27; 0.23]	[-0.29; 0.18]	[-0.18; 0.15]
36	-0.0205	-0.0376	-0.0311	-0.0101	-0.0443	-0.0158
	[-0.21; 0.16]	[-0.21; 0.12]	[-0.24; 0.17]	[-0.28; 0.24]	[-0.29; 0.18]	[-0.19; 0.14]
			Semiparametric multivariate restricted			
6			-0.0268	[-0.13; 0.08]		
12			-0.0259	[-0.12; 0.07]		
36			-0.0267	[-0.12; 0.07]		
			Maximum Likelihood			
univ.	-0.0569	-0.0795	-0.0596	-0.0367	-0.0676	-0.0604
unrestr.	-0.0121	-0.0123	-0.0121	-0.0121	-0.0121	-0.0114
restr.			-0.0122			

Table 3: Number of rejections for univariate diagnostic residuals tests of the VARMA(2,1) model (18) based on different transformations: the raw variance and covariance processes (vech), the Cholesky decomposition (chol), the symmetric matrix square root (sym-root), the matrix logarithm (mlog), the inverse as well as MBC and BC-DC given by (3) and (7), respectively, with estimated transformation parameters. Upper panel: Ljung-Box test with 10 lags for no autocorrelation in residuals. Middle panel: Ljung-Box tests with 10 lags for no autocorrelation in *squared* residuals. Lower panel: Jarque-Bera test for normality of the residuals.

P-Value	vech	chol	sym-root	mlog	inverse	MBC	BC-DC
			Residual autocorrelation				
0.05	21	14	14	8	15	7	6
0.01	21	9	11	4	11	4	2
0.001	17	6	10	1	7	1	1
			Conditional heteroskedasticity				
0.05	21	21	21	4	21	3	4
0.01	20	21	21	2	21	1	2
0.001	19	21	21	1	21	1	1
			Non-normality				
0.05	21	21	21	11	21	13	12
0.01	21	21	21	9	21	10	10
0.001	21	21	21	8	21	8	8

Table 4: Fraction of mean Frobenius loss (24) between bias-corrected (15) vs. naive (14) forecasts. The forecasts are from the diagonal VARMA(2,1) model (18) based on the MBC transform (3) and the BC-DC transform (7).

	$h = 1$		$h = 5$		$h = 10$		$h = 20$	
δ	MBC	BC-DC	MBC	BC-DC	MBC	BC-DC	MBC	BC-DC
-0.5	1.8150	4.0653	2.2496	2.1426	1.6510	1.7354	1.6164	1.3337
-0.1	0.8758	0.9068	0.8750	0.9083	0.8882	0.9120	0.8922	0.9043
-0.05	0.8847	0.9177	0.8920	0.9192	0.9031	0.9216	0.9052	0.9133
0	0.9037	0.9326	0.9098	0.9304	0.9172	0.9306	0.9166	0.9212
0.05	0.9286	0.9513	0.9317	0.9464	0.9357	0.9445	0.9336	0.9345
0.1	0.9375	0.9576	0.9393	0.9518	0.9414	0.9491	0.9378	0.9393
0.5	0.9956	0.9925	0.9893	0.9866	0.9853	0.9826	0.9792	0.9765
1	1.0011	1.3840	1.0021	1.3087	1.0026	1.2046	1.0029	1.1479

Table 5: Risks from VARMA(2,1) forecasts based on the MBC transform (3) and the BC-DC transform (7), as well as CAW(2,1) and CAW-DCC(2,1) benchmarks (models (20) and (21)-(23), respectively). Left: Frobenius loss function (24) from bias-corrected forecasts. Right: Negative logarithmic score (25) of density forecasts. Asterisks denote the 90% model confidence set for a given loss function and horizon h. The best-performing forecast is in boldface. Models for which at least one of the forecasts is not positive definite have missing predictive densities (—).

	Frobenius Loss				Predictive Density			
	$h = 1$	$h = 5$	$h = 10$	$h = 20$	$h = 1$	$h = 5$	$h = 10$	$h = 20$
MBC(-0.1)	92.45	138.98	176.74	211.33*	8.155	8.276*	8.356*	8.438*
MBC(-0.05)	92.53	139.75	177.42	212.45*	8.152	8.274*	8.351*	8.431*
MBC(0)	93.17	140.34	177.68	212.98*	**8.146***	**8.268***	**8.348***	**8.428***
MBC(0.5)	86.58*	141.55	179.39	217.35*	8.166	8.313	8.414	8.514
MBC(1)	89.50*	169.19	212.93	246.54*	—	—	—	—
BC-DC(-0.1)	84.86*	**133.85***	**170.98***	**206.42***	8.160	8.276*	8.355*	8.432*
BC-DC(-0.05)	**84.73***	134.01*	171.47*	207.29*	8.158	8.277*	8.359*	8.439*
BC-DC(0)	85.09*	134.43*	172.03*	208.07*	8.153	8.269*	8.352*	8.431*
BC-DC(0.5)	85.43*	142.84	181.56	218.82*	8.200	8.350	8.450	8.585
BC-DC(1)	122.39	218.40	258.76	284.48	—	—	—	—
Cholesky	86.20*	141.42	178.03	215.75*	8.179	8.318	8.415	8.519
CAW(2,1)	85.97*	137.34*	176.50	214.60*	8.187	8.318	8.410	8.503
CAW-DCC(2,1)	86.32*	137.77*	178.11	217.52*	8.302	8.415	8.497	8.616

Table 6: Risks from bias-corrected VARMA(2,1) forecasts based on the MBC transform (3) and the BC-DC transform (7), as well as CAW(2,1) and CAW-DCC(2,1) benchmarks (models (20) and (21)-(23), respectively). Left: Stein loss function (30). Right: L3 loss function (31). Asterisks denote the 90% model confidence set for a given loss function and horizon h. The best-performing forecast is in boldface. The case $\delta = 1$ is missing since there at least one of the forecasts is not positive-definite.

	Stein Loss				L3 Loss			
	$h = 1$	$h = 5$	$h = 10$	$h = 20$	$h = 1$	$h = 5$	$h = 10$	$h = 20$
MBC(-0.1)	0.997*	1.416*	1.720*	2.118*	201.0*	254.9*	293.9*	325.0*
MBC(-0.05)	0.997*	1.420	1.731*	2.142*	197.2*	254.4*	293.4*	324.3*
MBC(0)	0.996*	1.420	1.733*	2.149*	197.1*	254.9*	293.4*	324.4*
MBC(0.5)	1.076	1.573	1.936	2.430	183.0*	256.5*	294.5*	328.1*
BC-DC(-0.1)	1.000*	1.410*	1.697*	**2.074***	182.5*	249.6*	**288.3***	322.8*
BC-DC(-0.05)	0.996*	**1.404***	**1.696***	2.090*	181.9*	**249.5***	288.7*	**322.7***
BC-DC(0)	**0.995***	1.405*	1.698*	2.098*	182.3*	249.8*	289.0*	323.1*
BC-DC(0.5)	1.052	1.527	1.891	2.414	**181.7***	258.6	296.8*	327.9*
Cholesky	1.099	1.581	1.945	2.440	181.7*	256.6*	293.3*	327.3*
CAW(2,1)	1.025	1.436	1.740*	2.158*	184.3*	261.2	298.0*	332.6*
CAW-DCC(2,1)	1.004*	1.409*	1.712*	2.170*	186.6*	262.7	298.6	333.1*

Table 7: Risks from bias-corrected VARMA(2,1) forecasts based on the MBC transform (3) and the BC-DC transform (7), as well as CAW(2,1) and CAW-DCC(2,1) benchmarks (models (20) and (21)-(23), respectively). Left: Realized variance of minimum variance portfolio (32). Right: Squared daily return of minimum variance portfolio as defined below (32). Asterisks denote the 90% model confidence set for a given loss function and horizon h. The best-performing forecast is in boldface. The case $\delta = 1$ is missing since there at least one of the forecasts is not positive-definite.

	Realized variance of MVP				Squared daily return of MVP			
	$h = 1$	$h = 5$	$h = 10$	$h = 20$	$h = 1$	$h = 5$	$h = 10$	$h = 20$
MBC(-0.1)	0.7924	0.8036*	0.8107*	0.8253*	0.6414*	0.6550*	0.6703*	0.6906
MBC(-0.05)	0.7915*	0.8027*	0.8099*	0.8245*	0.6409*	0.6542*	0.6694*	0.6887
MBC(0)	**0.7912***	**0.8022***	0.8092*	0.8233*	0.6409*	0.6538*	0.6685*	0.6870*
MBC(0.5)	0.7920*	0.8030*	0.8119	0.8267*	0.6425*	0.6543*	0.6692*	0.6905
BC-DC(-0.1)	0.7915*	0.8025*	0.8089*	0.8230*	0.6396*	0.6520*	0.6663*	0.6842*
BC-DC(-0.05)	0.7919*	0.8030*	0.8098*	0.8243*	0.6399*	0.6520*	0.6667*	0.6856*
BC-DC(0)	0.7916*	0.8025*	**0.8088***	**0.8224***	0.6397*	0.6514*	0.6656*	**0.6840***
BC-DC(0.5)	0.7928	0.8034*	0.8121	0.8272*	0.6434*	0.6534*	0.6702*	0.6913
Cholesky	0.7921*	0.8026*	0.8109*	0.8249*	0.6409*	0.6535*	0.6665*	0.6896
CAW(2,1)	0.7930	0.8038*	0.8110	0.8241*	**0.6385***	0.6514*	**0.6630***	0.6891
CAW-DCC(2,1)	0.7928	0.8039*	0.8101*	0.8231*	0.6443*	**0.6496***	0.6632*	0.6869*

Figures

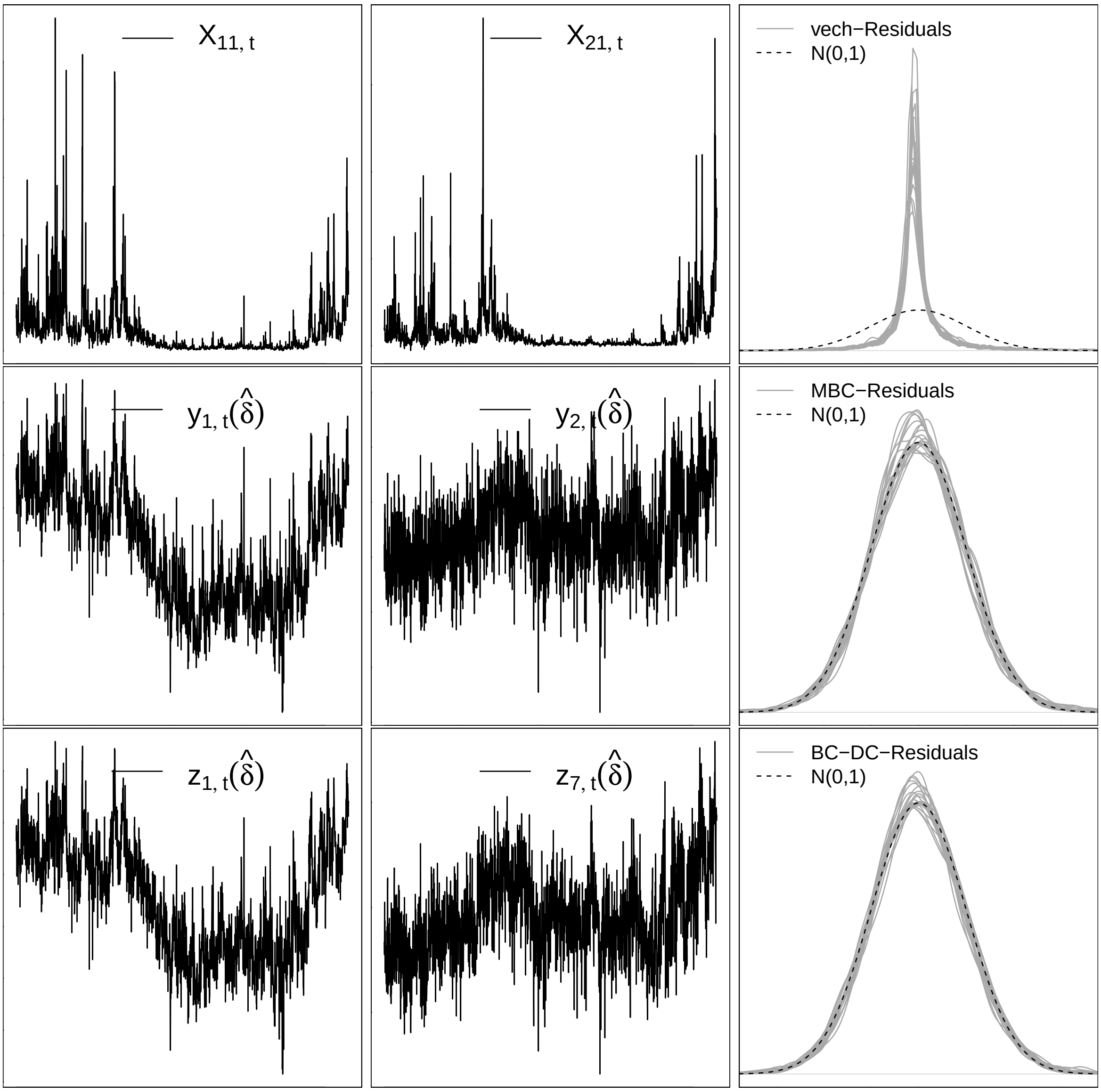

Figure 1: Left two panels: Time series plots of raw variance and covariance series (above), of the MBC-transformed data with δ estimated by the unrestricted semiparametric estimator (middle), as well as the BC-transformed variance $z_{1t}(\hat{\delta})$ and z-transformed correlation $z_{7t}(\hat{\delta}) = \tilde{R}_{12,t}$. Right panel: Kernel density estimates of standardized VARMA(2,1) residuals (grey) of the vech, MBC-RCov and BC-DC series. The standard normal density is shown as the black dashed line.

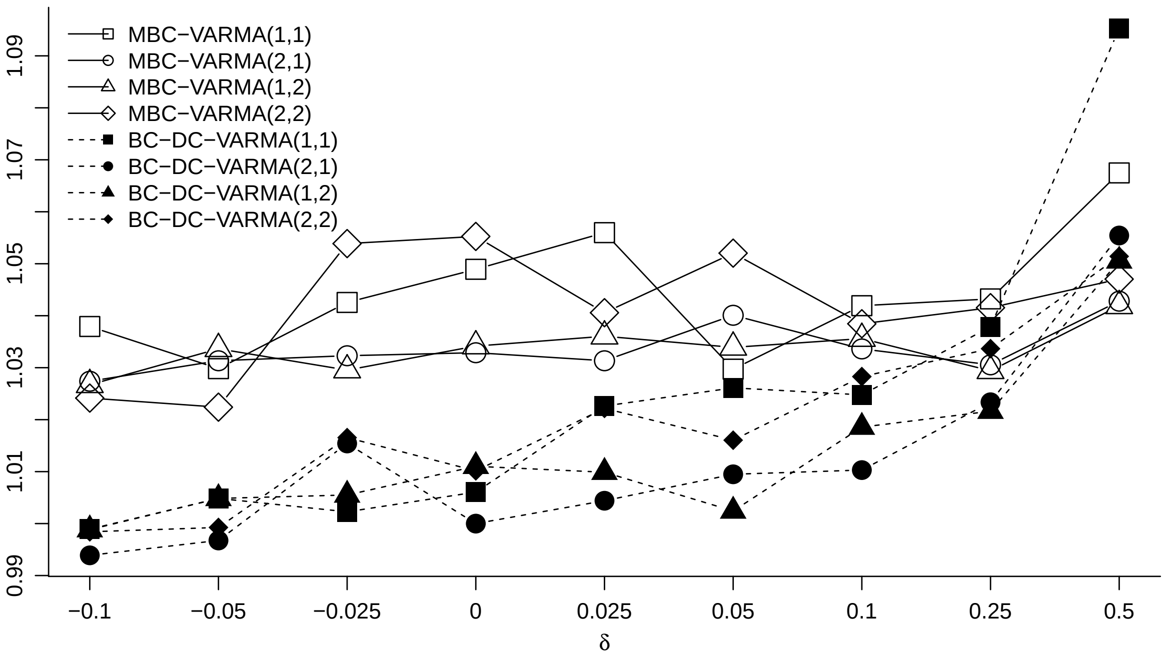

Figure 2: Robustness of out-of-sample results with respect to the order specification of the VARMA model (18). For $h = 10$, the Frobenius loss (24) is plotted as a fraction of the loss for the BC-DC-VARMA(2,1) specification with $\delta = 0$.

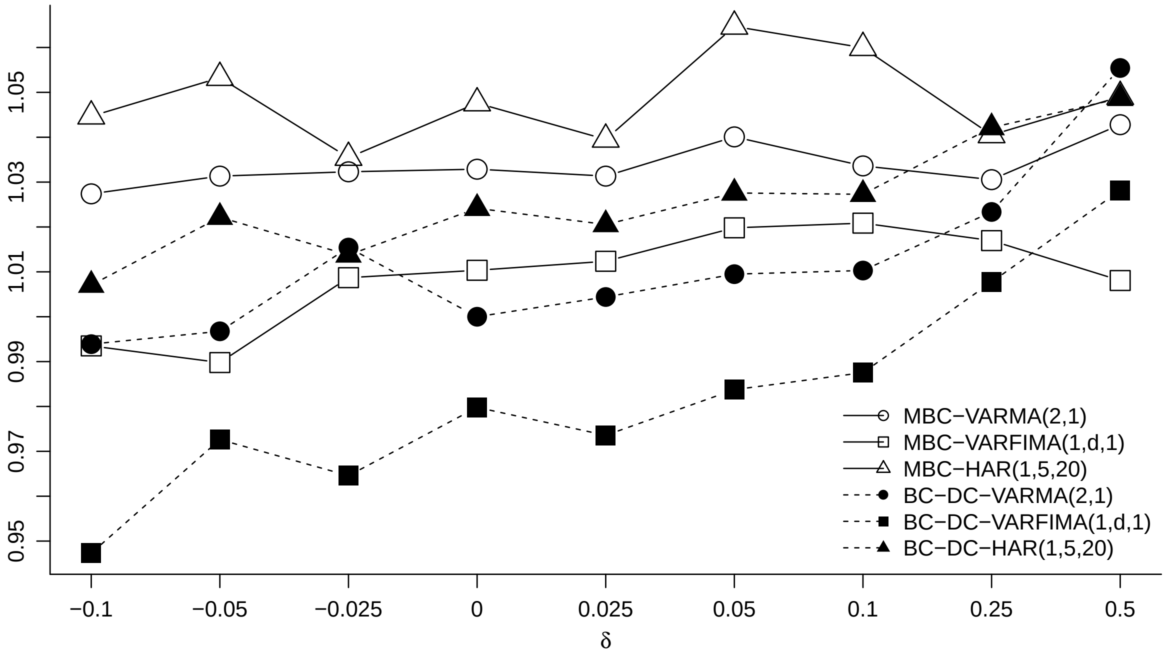

Figure 3: Robustness of out-of-sample results with respect to specification of dynamic persistence. The baseline VARMA(2,1) model (18) is compared to the VARFIMA model (27) and the HAR model (26). For $h = 10$, the Frobenius loss (24) is plotted as a fraction of the loss for the BC-DC VARMA(2,1) model with $\delta = 0$.

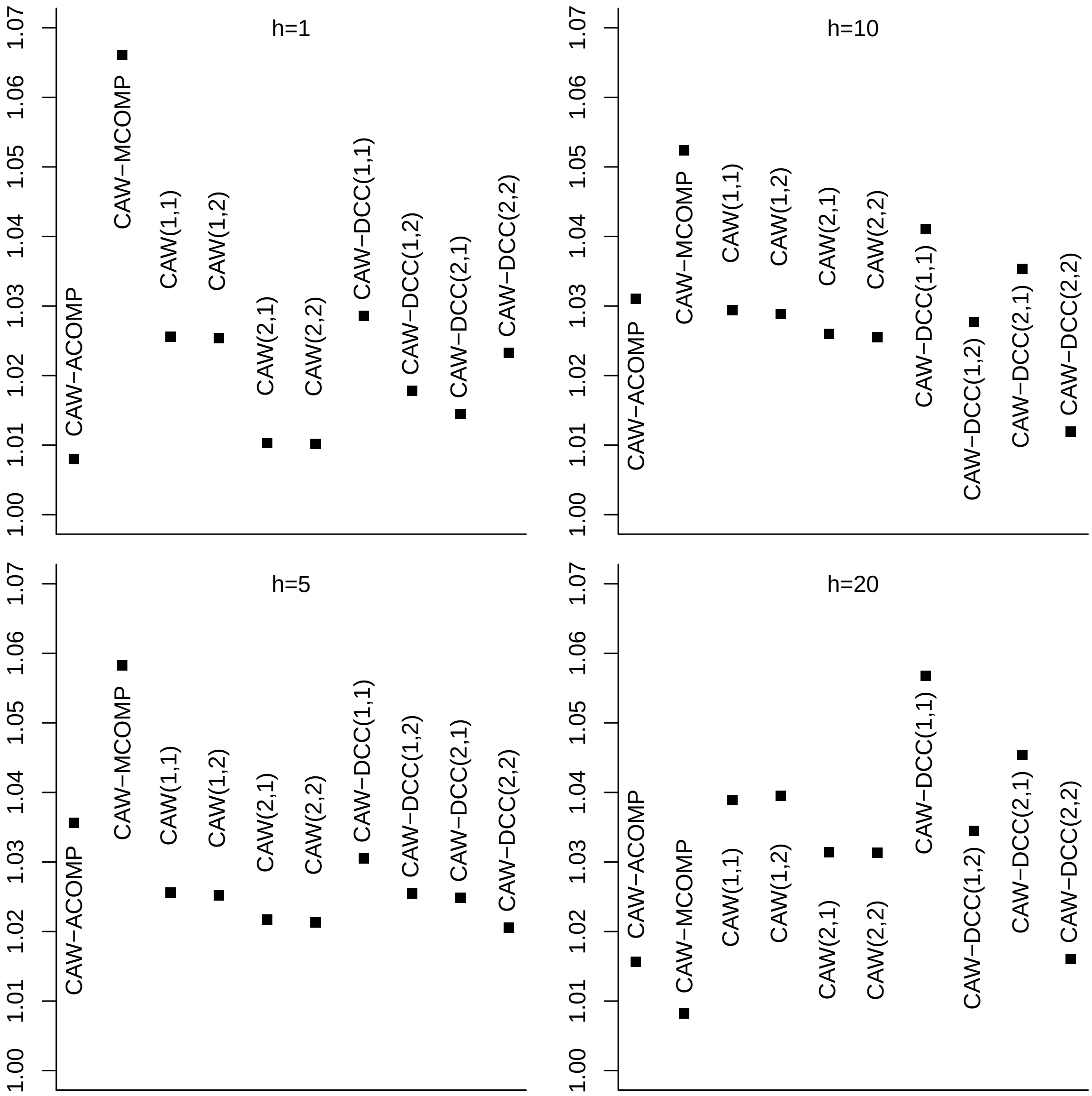

Figure 4: Robustness of out-of-sample results with respect to order specification of the CAW model (20) and the CAW-DCC model (21)-(23) and to the CAW-ACOMP model (28) and the CAW-MCOMP model (29). For each horizon, the Frobenius loss (24) is plotted as a fraction of the loss for the BC-DC-VARMA(2,1) model with $\delta = 0$.