A partial correlation vine based approach for modeling and forecasting multivariate volatility time-series

Nicole Barthel[*], Claudia Czado[*] and Yarema Okhrin[†]

February 18, 2019

Abstract

A novel approach for dynamic modeling and forecasting of realized covariance matrices is proposed. Realized variances and realized correlation matrices are jointly estimated. The one-to-one relationship between a positive definite correlation matrix and its associated set of partial correlations corresponding to any vine specification is used for data transformation. The model components therefore are realized variances as well as realized standard and partial correlations corresponding to a daily log-return series. As such, they have a clear practical interpretation. A method to select a regular vine structure, which allows for parsimonious time-series and dependence modeling of the model components, is introduced. Being algebraically independent the latter do not underlie any algebraic constraint. The proposed model approach is outlined in detail and motivated along with a real data example on six highly liquid stocks. The forecasting performance is evaluated both with respect to statistical precision and in the context of portfolio optimization. Comparisons with Cholesky decomposition based benchmark models support the excellent prediction ability of the proposed model approach.

Keywords: *Forecasting; Partial correlation vine; Realized volatility; Time-series modeling; R-vine structure selection*

1 Introduction

The increasing availability of high-frequency data makes volatility modeling and forecasting to one of the most vividly discussed topics in financial econometrics. Also, the strongly increasing interaction and interconnectedness between financial markets have stimulated the need for reliable modeling and forecasting techniques to capture the cross-sectional and temporal dependencies of financial asset returns. Especially during negative economic phases and periods of financial turmoil, assets become more dependent and linkages between asset market volatility tighten (Cappiello et al. 2006). This affects fields such as asset pricing, portfolio allocation and evaluation of risk.

High-frequency data allow to consistently estimate ex-post realized volatility and realized covariances using the sum of squared intra-day returns (Doléans-Dade and Meyer 1970; Jacod 1994). By making naturally latent variables, namely volatilities and covariances, observable and measurable, standard time-series approaches can be applied to model their realized counterparts. Building upon the aforementioned classical estimator first used in the context of high-frequency data by Barndorff-Nielsen and Shephard (2004), many refinements were investigated to improve its overall quality and precision (Zhang 2011), to reduce market

---

[*]Department of Mathematics, Technische Universität München, Germany (email: nicole.barthel@tum.de (corresponding author), cczado@ma.tum.de)

[†]Department of Statistics, Faculty of Business and Economics, Universität Augsburg, Germany (email: yarema.okhrin@wiwi.uni-augsburg.de)

microstructure noise (Gençay et al. 2001; Zhang et al. 2005) and to take into account jumps (Christensen et al. 2010) and asynchronicity (Hayashi et al. 2005).

The main modeling challenge when developing prediction tools for realized covariance matrices are the algebraic restrictions of symmetry and positive definiteness the forecasts need to satisfy. Direct modeling of the components using univariate time-series models does not meet this constraint (Andersen et al. 2006) and neglects e.g. dynamic volatility spillovers among the series of variances and covariances (Voev 2008). Several multivariate approaches such as the Wishart Autoregressive (WAR) model (Gouriéroux et al. 2009) and its dynamic counterpart the Conditional Autoregressive Wishart (CAW) model (Golosnoy et al. 2012) have been developed. Andersen et al. (2006) propose a multivariate generalization of the realized GARCH model (Hansen et al. 2012) by modifying the Dynamic Conditional Correlation (DCC) model of Engle (2002). The basic idea of the latter model is to split up the estimation problem into the two simpler tasks of modeling the conditional volatilities and the correlation dynamics. Halbleib and Voev (2014) adopt this strategy using high-frequency data in the volatility part and daily data in the correlation part at the expense of less flexible correlation specifications. As an alternative, data transformation is one of the most frequently used approaches. Bauer and Vorkink (2011) apply the matrix logarithm function and a factor model approach to the individual components, which, however, leads to a computationally demanding model. First proposed by Andersen et al. (2003) and having evolved to one of the standard ways to proceed, the Cholesky decomposition is a proven tool to guarantee symmetry and positive definiteness of the forecasts. For example, Chiriac and Voev (2011) decompose the series of realized covariance matrices via the Cholesky factorization and model the so-obtained series of Cholesky elements with a vector autoregressive fractionally integrated moving average (VARFIMA) process. Brechmann et al. (2018) build upon this model approach, but pay special attention to the specific dependencies among the Cholesky series induced by the nonlinear data transformation. While Cholesky decomposition based models are straightforward and easy to implement, they also come with drawbacks. There is no clear interpretation of the model components obtained after data transformation and the latter induces an additive bias in the forecasts of the original data due to its nonlinear nature. Also, the Cholesky decomposition depends on the ordering of the data within the realized covariance matrices with no obvious way to fix the order in advance. Complete enumeration leads to a computationally expensive estimation problem. On the other hand, fixing the order upfront ignores a possible changing behavior of the data over time.

Irrespective of the considered data transformation, multivariate approaches for time-series modeling often suffer from lacking flexibility in the parameters. Further, they barely allow for convenient modeling of non-Gaussianity and conditional heteroscedasticity, which, however, are typical features of volatility data. In comparison, univariate time-series models allow for various extensions and refinements to tackle these problems. Besides ARFIMA processes (Andersen et al. 2006), heterogeneous autoregressive (HAR) processes are most commonly applied to (log-transformed) realized volatility time-series capturing their long-memory behavior. They include volatility measured over different time horizons and account for multifractal scaling (Corsi 2009). Both ARFIMA and HAR models can be extended by e.g. GARCH augmentations to account for non-Gaussianity and volatility clustering (Corsi et al. 2008). By considering skewed error distributions for the residuals, typically observed high skewness and kurtosis can be additionally captured (Bai et al. 2003; Fernández and Steel 1998).

In the light of the above discussion, a tool to transform the realized covariance matrices, which allows for reasonable computational effort, interpretability of the model components obtained after data transformation and to exploit the beneficial features of univariate time-series modeling, is desirable. A promising candidate which meets these requirements are partial correlation vines. The latter assign partial correlations to the edges of an R-vine tree structure. The latter is a graph theoretical object first proposed by Bedford and Cooke

(2002), which consists of a set of linked trees specifying bivariate conditional constraints. The set of standard and partial correlations specified through an R-vine structure has attractive properties. Bedford and Cooke (2002) proof that there is a bijection between the specified (partial) correlations and the set of symmetric and positive definite correlation matrices. Further, Kurowicka and Cooke (2003) find that any partial correlation vine specifies algebraically independent (partial) correlations, i.e. the latter can take arbitrary values in $(-1, 1)$ while still guaranteeing positive definiteness of the corresponding correlation matrix. This result advocates partial correlation vines to be a useful tool in several applications. Kurowicka and Cooke (2006) use them to solve the completion problem for positive definite matrices, whereas Lewandowski et al. (2009) introduce a method to uniformly generate random correlation matrices from the space of positive definite correlation matrices. Brechmann and Joe (2014) base a parsimonious parameterization of correlation matrices on partial correlation vines in combination with factor analysis and Brechmann and Joe (2015) use these findings to capture the dependence structure in multivariate data. Considering financial data, Poignard (2017) introduces a vine-GARCH approach as flexible multivariate GARCH-type model, which parametrizes the latent correlations appearing in the DCC model of Engle (2002) in terms of a partial correlation vine. Based on the specific nature of an R-vine tree structure, their estimation technique proceeds iteratively by evoking only bivariate GARCH models in each tree level and thus allows for dimension reduction as compared to computationally highly demanding classical multivariate GARCH models.

To our knowledge, data transformation using partial correlation vines has not yet been investigated to model and forecast multivariate realized volatility time-series. We propose a joint estimation and prediction model of the realized variance times-series and a subset of realized standard and partial correlation time-series. The latter are obtained after transforming the series of realized correlation matrices based on an R-vine structure as first step of the model approach. To select among the large number of possible R-vine structures the one used for data transformation, we propose a selection method, which exclusively relies on historical information of the modeled time-series and thus dynamically adapts to changing data behavior over time. We will show that data transformation based on this R-vine structure further allows for parsimony in the resulting multivariate time-series models, which are to be estimated as second step of the model approach. We opt for a copula based time-series model to exploit the beneficial features of elaborate univariate time-series models. By considering flexible copulas for the dependence between the model components possible asymmetry and nonlinearity can be captured. Combining in a third step the predicted realized variances and the predicted realized correlation matrix obtained after back-transformation of the underlying realized partial correlation vine guarantees a symmetric and positive definite realized covariance matrix forecast.

The paper is structured as follows. In Section 2, we introduce partial correlation vines combining the notion of partial correlations and an R-vine structure. The transformation of a correlation matrix to a partial correlation vine based on a given R-vine structure and vice versa is explained in detail. In Section 3, we introduce the general data setting and motivate the choice of Cholesky decomposition based models as our main benchmarks. In Section 4, we outline in detail the three main steps of the proposed partial correlation vine data transformation approach including R-vine structure selection in Section 4.2 and multivariate time-series modeling in Section 4.3. Supported by the analysis of high-frequency data for six stocks listed on the NYSE, AMEX and NASDAQ beneficial properties of the proposed modeling strategy will be explored. In Section 5, detailed investigation of the real data example will be continued. Section 5.4 shows the excellent forecasting performance of the partial correlation vine data transformation approach both with respect to statistical precision and mean-variance balance in portfolio optimization. This paper comes with extensive supplementary material.

2   Partial correlation vines

First, we provide necessary background on the two main ingredients of the proposed model approach – partial correlations and regular vines.

2.1   Partial correlations

We consider a random vector $\boldsymbol{X}_{\mathcal{I}} := (X_1, \ldots, X_d)$, $d \geq 2$, with zero mean, where $\mathcal{I}$ is the index set $\{1, \ldots, d\}$. We denote the $d \times d$ covariance matrix by $\boldsymbol{Y}$ and obtain the corresponding $d \times d$ correlation matrix $\boldsymbol{R}$ as $\boldsymbol{R} = \boldsymbol{D}^{-1/2} \boldsymbol{Y} \boldsymbol{D}^{-1/2}$, where $\boldsymbol{D} = \mathrm{diag}\,(y_{1,1}, \ldots, y_{d,d})$ is the diagonal matrix of variances. Further, we define a subset $L \subseteq \mathcal{I}$ having at least cardinality 2, i.e. $|L| \geq 2$. For a pair $i, j \in L, i \neq j$, we denote L with the subset $\{i, j\}$ removed by $D_{\{i,j\}} := L_{-\{i,j\}} = L \backslash \{i, j\}$ and the corresponding random vector by $\boldsymbol{X}_{D_{\{i,j\}}} := \{X_k, k \in D_{\{i,j\}}\}$. The partial regression coefficients $b_{i,j;D_{\{i,j\}}}$ are defined as the quantities that minimize

$$\mathbb{E}\Big[(X_i - \sum_{j \in L_{-\{i\}}} b_{i,j;D_{\{i,j\}}} X_j)^2\Big].$$

The corresponding partial correlation coefficients $\rho_{i,j;D_{\{i,j\}}}$ quantify the dependence between X_i and X_j without the linear effect of $\boldsymbol{X}_{D_{\{i,j\}}}$ and are defined by (Kurowicka and Joe 2011, p. 47)

$$\rho_{i,j;D_{\{i,j\}}} := \mathrm{sgn}(b_{i,j;D_{\{i,j\}}}) \left(b_{i,j;D_{\{i,j\}}} b_{j,i;D_{\{i,j\}}}\right)^{1/2}.$$

We refer to the cardinality of $D_{\{i,j\}}$ as order of the partial correlation coefficient. For order zero, i.e. $|L| = |\{i, j\}| = 2$ and thus $D_{\{i,j\}} = \emptyset$, we obtain pairwise standard correlations between X_i and X_j, $i, j \in \mathcal{I}, i \neq j$. We write $\rho_{i,j;\emptyset} = \rho_{i,j}$. Now, consider for a subset $L \subseteq \mathcal{I}$ of at least cardinality 3, a set of distinct indices $\{i, j, k\} \subseteq L$, $i \neq j \neq k$. We define $\tilde{D} := L_{-\{i,j,k\}}$ such that $D_{\{i,j\}} = \tilde{D} \cup k$. Anderson (1958) derives a formula to recursively calculate the partial correlations of any order $|D_{\{i,j\}}|$ with $|D_{\{i,j\}}| \geq 1$ in terms of (partial) correlations of lower order. With $\rho^2_{i,k;\tilde{D}} < 1$ and $\rho^2_{j,k;\tilde{D}} < 1$ it holds that

$$\rho_{i,j;D_{\{i,j\}}} = \frac{\rho_{i,j;\tilde{D}} - \rho_{i,k;\tilde{D}}\rho_{j,k;\tilde{D}}}{\sqrt{1 - \rho^2_{i,k;\tilde{D}}}\sqrt{1 - \rho^2_{j,k;\tilde{D}}}}. \tag{1}$$

Since the evaluation of higher order partial correlations gets too involved when exclusively relying on this recursion formula, in practice typically a more efficient calculation procedure is used (see e.g. Whittaker (2009)). Let $\boldsymbol{\Omega}$ be the submatrix of standard correlations with indices $L \subseteq \mathcal{I}$, i.e. $\boldsymbol{\Omega} = (\omega_{k,\ell})_{k,\ell=1,\ldots,|L|} = (\rho_{l_k l_\ell})_{k,\ell=1,\ldots,|L|}$, where l_k is the k-th element in L. Let $\boldsymbol{P}$ be its inverse, i.e. $\boldsymbol{P} = \boldsymbol{\Omega}^{-1} = (p_{k,\ell})_{k,\ell=1,\ldots,|L|}$. Then, it holds

$$\rho_{l_k,l_\ell;D_{\{l_k,l_\ell\}}} = -\frac{p_{k,\ell}}{\sqrt{p_{k,k} p_{\ell,\ell}}}. \tag{2}$$

Thus, through inversion of $\boldsymbol{\Omega}$ all partial correlations between X_i and X_j $(i, j \in L, i \neq j)$ given all other variables $\boldsymbol{X}_{D_{\{i,j\}}}$ are simultaneously calculated. If interest is in a single partial correlation $\rho_{i,j;D_{\{i,j\}}}$ for $i, j \in L$ with $i \neq j$ fixed, computing complexity can be reduced by assorting $\boldsymbol{\Omega}$ blockwise with indices (i, j) and $D_{\{i,j\}}$, i.e.

$$\boldsymbol{\Omega}^{-1} = \begin{pmatrix} \boldsymbol{\Omega}_{1,1} & \boldsymbol{\Omega}_{1,2} \\ \boldsymbol{\Omega}_{2,1} & \boldsymbol{\Omega}_{2,2} \end{pmatrix}^{-1} = \boldsymbol{P} = \begin{pmatrix} \boldsymbol{P}_{1,1} & \boldsymbol{P}_{1,2} \\ \boldsymbol{P}_{2,1} & \boldsymbol{P}_{2,2} \end{pmatrix},$$

where $\boldsymbol{\Omega}_{1,1}$ is a 2×2 matrix with elements $\omega_{1,1} = \omega_{2,2} = 1$, $\omega_{1,2} = \omega_{2,1} = \rho_{i,j}$ and $\boldsymbol{P}_{1,1}$ its counterpart with elements $p_{1,1}$, $p_{1,2} = p_{2,1}$, $p_{2,2}$. Using standard results for block matrix inversion (see e.g. Bernstein (2005)) we have $\boldsymbol{P}_{1,1}^{-1} = \boldsymbol{\Omega}_{1,1} - \boldsymbol{\Omega}_{1,2}\boldsymbol{\Omega}_{2,2}^{-1}\boldsymbol{\Omega}_{2,1}$ with elements $\tilde{p}_{1,1}$, $\tilde{p}_{1,2} = \tilde{p}_{2,1}$, $\tilde{p}_{2,2}$. We conclude that

$$\rho_{i,j;D_{\{i,j\}}} \overset{(2)}{=} -\frac{p_{1,2}}{\sqrt{p_{1,1}p_{2,2}}} = -\frac{-\frac{1}{\det \boldsymbol{P}_{1,1}}\tilde{p}_{1,2}}{\sqrt{\frac{1}{\det \boldsymbol{P}_{1,1}}\tilde{p}_{1,1}\frac{1}{\det \boldsymbol{P}_{1,1}}\tilde{p}_{2,2}}} = \frac{\tilde{p}_{1,2}}{\sqrt{\tilde{p}_{1,1}\tilde{p}_{2,2}}}. \tag{3}$$

From now on, we refer to $\mathcal{C}_d$ as the set of all standard correlations and to $\mathcal{C}_d^{\mathrm{p}}$ as the set of all pairwise standard and partial correlations. The $1 \times \binom{d}{2}$ vector $\boldsymbol{P}_{\mathcal{C}_d}$ and the $1 \times \binom{d}{2}2^{d-2}$ vector $\boldsymbol{P}_{\mathcal{C}_d^{\mathrm{p}}}$ record all standard correlations and all standard and partial correlations, respectively, of the random vector $\boldsymbol{X}_{\mathcal{I}}$ in lexicographical order with increasing subset $L \subseteq \mathcal{I}$, i.e.

$$\boldsymbol{P}_{\mathcal{C}_d} := \left(\rho_{1,2}, \ldots, \rho_{1,d}, \rho_{2,3}, \ldots, \rho_{2,d}, \ldots, \rho_{(d-1),d}\right)$$

and

$$\boldsymbol{P}_{\mathcal{C}_d^{\mathrm{p}}} := \big(\boldsymbol{P}_{\mathcal{C}_d},$$
$$\rho_{1,2;3}, \ldots, \rho_{1,2;d}, \rho_{1,3;2}, \ldots, \rho_{1,d;(d-1)}, \rho_{2,3;1}, \ldots, \rho_{(d-1),d;(d-2)},$$
$$\ldots,$$
$$\rho_{1,2;3,\ldots,d}, \ldots, \rho_{1,d;2,\ldots,(d-1)}, \ldots, \rho_{(d-1),d;1,\ldots,(d-2)}\big).$$

To conclude and as illustrated in Figure 1, from a $d \times d$ covariance matrix $\boldsymbol{Y}$ the $1 \times d$ vector of variances $\boldsymbol{y}$ and the $d \times d$ correlation matrix $\boldsymbol{R}$ can be obtained. The latter fully determines the vector $\boldsymbol{P}_{\mathcal{C}_d^{\mathrm{p}}}$, which takes values in $(-1, 1)$ and collects all $\binom{d}{2}2^{d-2}$ standard and partial correlations. In the following, we will show that the other way round the correlation matrix $\boldsymbol{R}$ can be uniquely determined from only a few elements of $\boldsymbol{P}_{\mathcal{C}_d^{\mathrm{p}}}$, which are defined through a regular vine.

Data: random vector $\boldsymbol{X}_{\mathcal{I}} \in \mathbb{R}^d$
with covariance matrix $\boldsymbol{Y} \in \mathbb{R}^{d \times d}$

Variance vector of $\boldsymbol{X}_{\mathcal{I}}$
$\boldsymbol{y}' = (y_{1,1}, \ldots, y_{d,d})' \in \mathbb{R}_{>0}^d$

Correlation matrix of $\boldsymbol{X}_{\mathcal{I}}$
$$\boldsymbol{R} = \begin{pmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,d} \\ \rho_{1,2} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,d} & \cdots & \cdots & 1 \end{pmatrix} \in (-1,1)^{d \times d}$$

Vector of standard correlations of $\boldsymbol{X}_{\mathcal{I}}$
$\boldsymbol{P}'_{\mathcal{C}_d} = \left(\rho_{1,2}, \ldots, \rho_{(d-1),d}\right)' \in (-1,1)^{\binom{d}{2}}$

formulas (1), (2)

Vector of standard and partial correlations of $\boldsymbol{X}_{\mathcal{I}}$
$\boldsymbol{P}'_{\mathcal{C}_d^{\mathrm{p}}} = \left(\rho_{1,2}, \ldots, \rho_{(d-1),d;1,\ldots,(d-2)}\right)' \in (-1,1)^{\binom{d}{2}2^{d-2}}$
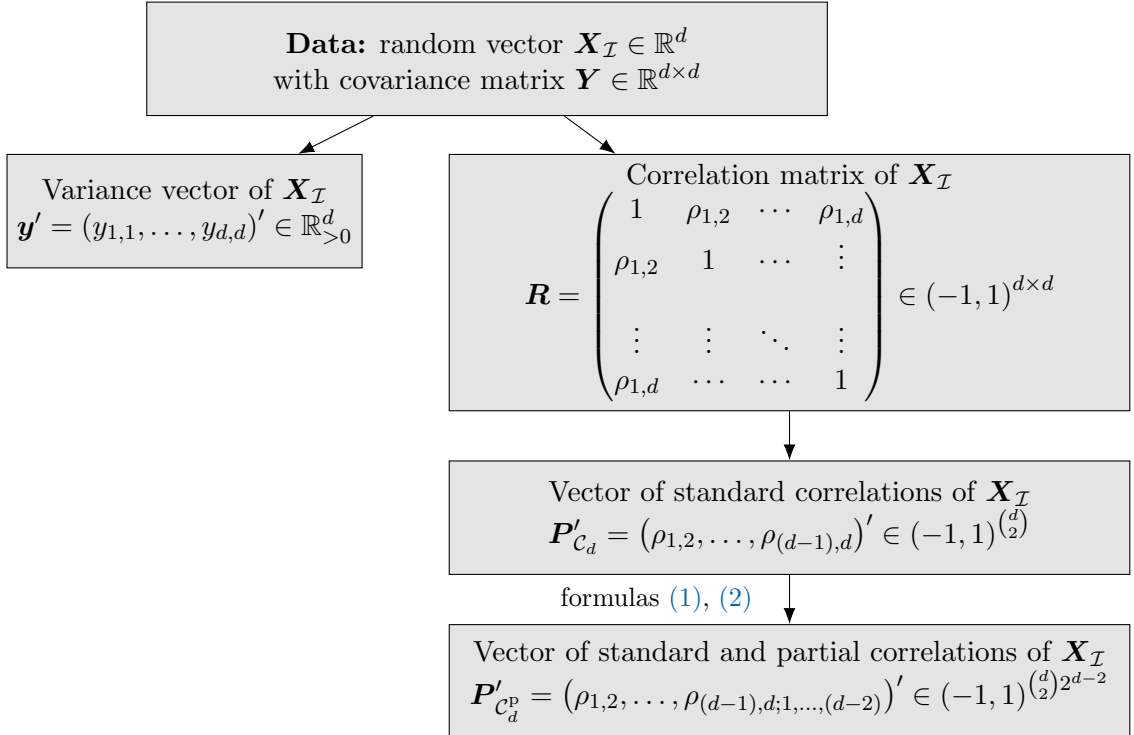
Figure 1: Data prespecified by a given covariance matrix $\boldsymbol{Y}$.

2.2 Regular vines

According to Bedford and Cooke (2002), a regular vine (R-vine) on d elements is a set of $d - 1$ linked trees, i.e. undirected and acyclic graphs, $\mathcal{V}_d := (\mathcal{T}_1, \ldots, \mathcal{T}_{d-1})$ with the set of edges $E(\mathcal{V}_d) := E_1 \cup \cdots \cup E_{d-1}$ and the set of nodes $N(\mathcal{V}_d) := N_1 \cup \cdots \cup N_{d-1}$ such that

(i) $\mathcal{T}_1$ is a tree with nodes $N_1 = \{1, \ldots, d\}$ and edges E_1,

(ii) for $\ell = 2, \ldots, d - 1$, $\mathcal{T}_\ell$ is a tree with nodes $N_\ell = E_{\ell-1}$ and edges E_ℓ,

(iii) the *proximity condition* holds: For $\ell = 2, \ldots, d - 1$, whenever two nodes of $\mathcal{T}_\ell$ are connected by an edge, the corresponding edges of $\mathcal{T}_{\ell-1}$ share a node.

According to property (ii), the $d - (\ell - 1)$ edges $E_{\ell-1}$ in $\mathcal{T}_{\ell-1}$ become nodes in $\mathcal{T}_\ell$. Based on this linkage, each sequence of trees of an R-vine – from now on referred to as R-vine structure – allows to identify a set of $\binom{d}{2}$ (conditional) bivariate constraints. We refer to Kurowicka and Cooke (2003) and consider an arbitrary edge $e = \{a, b\} \in E_\ell$ of $\mathcal{V}_d$, $2 \le \ell \le d - 1$, with $a, b \in N_\ell$. Its *complete union* U_e^* is the subset of nodes in $\mathcal{T}_1$, i.e. the subset of $\{1, \ldots, d\}$, reachable from e by the membership relation, i.e.

$$U_e^* := \{n \in N_1 : \exists e_1 \in E_1, \ldots, e_{\ell-1} \in E_{\ell-1} : n \in e_1 \in \cdots \in e_{\ell-1} \in e\}.$$

The *conditioning set* D_e corresponding to $e = \{a, b\}$ is the intersection of the complete unions U_a^* and U_b^* corresponding to the edges $a, b \in E_{\ell-1}$, i.e.

$$D_e := U_a^* \cap U_b^*.$$

The corresponding symmetric difference is referred to as *conditioned set*

$$\{C_{e,a}, C_{e,b}\} := \{U_a^* \backslash D_e, U_b^* \backslash D_e\}.$$

By definition, each conditioned set in $\mathcal{V}_d$ consists of two single elements and forms a unique pair of variables $\{i, j\}$ with $i, j \in \{1, \ldots, d\}$, $i \ne j$. Thus, each pair is modeled by $\mathcal{V}_d$ exactly once either unconditioned, if it appears in the first tree level, or via conditioning if it appears in tree levels $\ell = 2, \ldots, d - 1$. If $\mathcal{V}_d$ specifies in each tree level one central node being attached to all edges, we speak of a C-vine. The latter reflects an ordering by importance.

Example 2.1. Figure 2 shows an R-vine structure on six elements labeled with the conditioned set and the conditioning set corresponding to each edge. The latter is indicated by a leading "|". The bold tree segment in $\mathcal{T}_2$ corresponds to the edge $e = \{\{1, 2\}, \{2, 6\}\}$. Reachable from edge $\{1, 2\} \in \mathcal{T}_1$ and $\{2, 6\} \in \mathcal{T}_1$ are the nodes $1, 2 \in N_1$ and $2, 6 \in N_1$, respectively. Thus, $D_e = U_{\{1,2\}}^* \cap U_{\{2,6\}}^* = \{1, 2\} \cap \{2, 6\} = \{2\}$ is the conditioning set corresponding to e and the conditioned set is $\{C_{e,\{1,2\}}, C_{e,\{2,6\}}\} = \{\{1, 2\} \backslash \{2\}, \{2, 6\} \backslash \{2\}\} = \{1, 6\}$.

2.3 Partial correlation vines

Now, building the bridge between partial correlations determined by a $d \times d$ correlation matrix and an R-vine structure $\mathcal{V}_d$ is straightforward: in a partial correlation vine with R-vine structure $\mathcal{V}_d$ each edge $e = \{a, b\} \in E(\mathcal{V}_d)$ is identified with the partial correlation coefficient $\rho_{C_{e,a}, C_{e,b}; D_e}$ that coincides with the conditioned and conditioning set specified by e. Thus, to each edge in $\mathcal{V}_d$ a value in $(-1, 1)$ is assigned. We define the set of the $\binom{d}{2}$ standard and partial correlations specified by $\mathcal{V}_d$ as $\mathcal{C}(\mathcal{V}_d)$ and denote by $\boldsymbol{P}_{\mathcal{C}(\mathcal{V}_d)}$ the $1 \times \binom{d}{2}$ vector
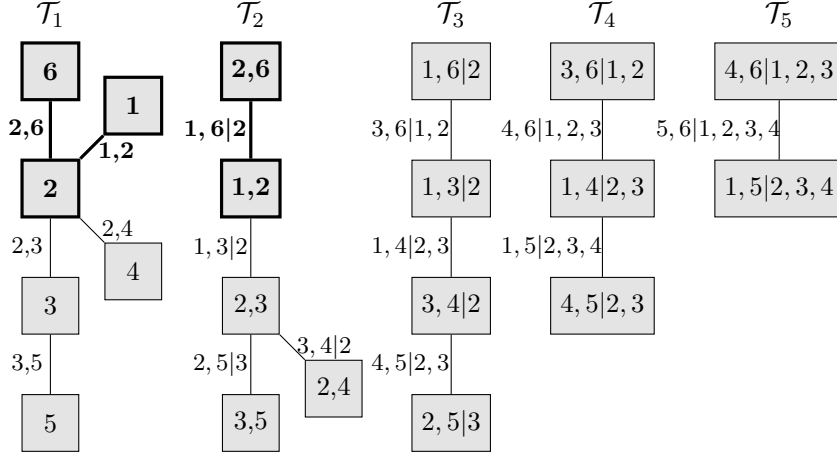
Figure 2: Example of a 6-dimensional R-vine structure with conditioning and conditioned sets corresponding to each edge.

that collects the corresponding values specified by the correlation matrix $\boldsymbol{R}$ in lexicographical order.

Bedford and Cooke (2002) provide the fundamental result that for any R-vine structure $\mathcal{V}_d$ there is a one-to-one relationship between the set of $d \times d$ positive definite correlation matrices and its set $\mathcal{C}(\mathcal{V}_d)$, i.e. for each R-vine structure $\mathcal{V}_d$ there exists a bijection

$$F_{\text{Cor2PCor}} : (-1, 1)^{\binom{d}{2}} \to (-1, 1)^{\binom{d}{2}}, \ F_{\text{Cor2PCor}}\left(\boldsymbol{P}_{\mathcal{C}_d}\right) = \boldsymbol{P}_{\mathcal{C}(\mathcal{V}_d)}. \tag{4}$$

In particular, according to Kurowicka and Cooke (2003) the elements in $\boldsymbol{P}_{\mathcal{C}(\mathcal{V}_d)}$ are algebraically independent, i.e. for any arbitrary assignment of values in $(-1, 1)$ to the edges of R-vine structure $\mathcal{V}_d$ the correlation matrix calculated from $\boldsymbol{P}_{\mathcal{C}(\mathcal{V}_d)}$ using the 'inverse' of (4) is positive definite with correlation values in $(-1, 1)$ for all off-diagonal elements. An efficient implementation of the bijection F_{Cor2PCor} and its 'inverse' is available in the R-package `VineCopula` (Schepsmeier et al. 2017). Pseudo-code is provided in Joe (2014). Note that while in the derivation of (3) and in the following explanations we assume the submatrix of standard correlations $\boldsymbol{\Omega}$ to be assorted blockwise with indices (i, j) and $D_{\{i,j\}}$, Joe (2014) assorts the indices using the order $D_{\{i,j\}}$ and (i, j).

Example 2.1 (continued)**.** We illustrate the data transformation based on R-vine structure $\mathcal{V}_6$ in Figure 2. As illustrated below, each standard correlation in $\boldsymbol{R}$ on the left-hand side is specified in the partial correlation vine corresponding to $\mathcal{V}_6$ through a (partial) correlation of order $\ell - 1$ modeled in tree $\mathcal{T}_\ell$ $(\ell = 1, \ldots, 5)$, i.e.

$$\boldsymbol{R} = \begin{pmatrix} \rho_{1,2} & \rho_{1,3} & \rho_{1,4} & \rho_{1,5} & \rho_{1,6} \\ & \rho_{2,3} & \rho_{2,4} & \rho_{2,5} & \rho_{2,6} \\ & & \rho_{3,4} & \rho_{3,5} & \rho_{3,6} \\ & & & \rho_{4,5} & \rho_{4,6} \\ & & & & \rho_{5,6} \end{pmatrix} \rightleftarrows \begin{pmatrix} \rho_{1,2} & \rho_{1,3;2} & \rho_{1,4;2,3} & \rho_{1,5;2,3,4} & \rho_{1,6;2} \\ & \rho_{2,3} & \rho_{2,4} & \rho_{2,5;3} & \rho_{2,6} \\ & & \rho_{3,4;2} & \rho_{3,5} & \rho_{3,6;1,2} \\ & & & \rho_{4,5;2,3} & \rho_{4,6;1,2,3} \\ & & & & \rho_{5,6;1,2,3,4} \end{pmatrix}$$

Transformation of correlation matrix. First, we derive from $\boldsymbol{R}$ the partial correlations corresponding to $\mathcal{V}_6$. Thus, proceeding in the illustration of the above matrices is from left to right. While the standard correlations in $\mathcal{T}_1$ can simply be taken from the correlation matrix $\boldsymbol{R}$, the first order partial correlations in $\mathcal{T}_2$ can be calculated using recursion formula (1), e.g.

$$\rho_{1,6;2} = \frac{\rho_{1,6} - \rho_{1,2}\rho_{2,6}}{\sqrt{1 - \rho_{1,2}^2}\sqrt{1 - \rho_{2,6}^2}}.$$

From tree level $\ell = 3$ on, we rely on formula (3) and elementwise calculate the partial correlations specified by $\mathcal{T}_3$ to $\mathcal{T}_5$. For example, for $\rho_{3,6;1,2}$ we set

$$\boldsymbol{\Omega}_{1,1} = \begin{pmatrix} 1 & \rho_{3,6} \\ \rho_{3,6} & 1 \end{pmatrix}, \ \boldsymbol{\Omega}_{1,2} = \begin{pmatrix} \rho_{1,3} & \rho_{2,3} \\ \rho_{1,6} & \rho_{2,6} \end{pmatrix}, \ \boldsymbol{\Omega}_{2,1} = \begin{pmatrix} \rho_{1,3} & \rho_{1,6} \\ \rho_{2,3} & \rho_{2,6} \end{pmatrix} \ \text{and} \ \boldsymbol{\Omega}_{2,2} = \begin{pmatrix} 1 & \rho_{1,2} \\ \rho_{1,2} & 1 \end{pmatrix}$$

and evaluate $\begin{pmatrix} \tilde{p}_{1,1} & \tilde{p}_{1,2} \\ \tilde{p}_{1,2} & \tilde{p}_{2,2} \end{pmatrix} = \boldsymbol{\Omega}_{1,1} - \boldsymbol{\Omega}_{1,2}\boldsymbol{\Omega}_{2,2}^{-1}\boldsymbol{\Omega}_{2,1}$. Then, we calculate

$$\rho_{3,6;1,2} \stackrel{(3)}{=} \frac{\tilde{p}_{1,2}}{\sqrt{\tilde{p}_{1,1}\tilde{p}_{2,2}}}.$$

Back-transformation to correlation matrix. Now, proceeding in the illustration of the above matrices is from right to left. We proceed treewise. The standard correlations from $\mathcal{T}_1$ can directly be taken. To calculate the standard correlations that correspond to the conditioned sets of the first order partial correlations available in $\mathcal{T}_2$, we use recursion formula (1), e.g.

$$\rho_{1,6} = \rho_{1,6;2}\sqrt{1 - \rho_{1,2}^2}\sqrt{1 - \rho_{2,6}^2} + \rho_{1,2}\rho_{2,6}.$$

Note that due to the proximity condition of an R-vine structure all standard correlations needed for this evaluation are available from the previous step.

From tree level $\ell = 3$ on, we rely on formula (3) to calculate the standard correlations that correspond to the conditioned sets of the partial correlations specified in $\mathcal{T}_3$ to $\mathcal{T}_5$. For example, to obtain $\rho_{3,6}$ we set $\boldsymbol{\Omega}_{1,2}$, $\boldsymbol{\Omega}_{2,2}$ and $\boldsymbol{\Omega}_{2,1}$ as above. Due to the proximity condition all standard correlations to do so are available from previous steps $\ell = 1, 2$. We calculate

$$\begin{pmatrix} q_{1,1} & q_{1,2} \\ q_{1,2} & q_{2,2} \end{pmatrix} = \boldsymbol{\Omega}_{1,2}\boldsymbol{\Omega}_{2,2}^{-1}\boldsymbol{\Omega}_{2,1}$$

such that $\tilde{p}_{1,1} = 1 - q_{1,1}$, $\tilde{p}_{1,2} = \tilde{p}_{2,1} = \rho_{3,6} - q_{1,2}$, $\tilde{p}_{2,2} = 1 - q_{2,2}$ and obtain

$$\rho_{3,6} \stackrel{(3)}{=} \rho_{3,6;1,2}\sqrt{(1 - q_{1,1})(1 - q_{2,2})} + q_{1,2}.$$

To conclude, the set of all standard correlations $\mathcal{C}_d$ can be determined from any set $\mathcal{C}(\mathcal{V}_d)$ specified by the partial correlation vine with R-vine structure $\mathcal{V}_d$. In particular, positive definiteness of the correlation matrix is always guaranteed. Figure 3 provides a summary overview of the relationships between the sets $\mathcal{C}_d$, $\mathcal{C}_d^{\mathrm{p}}$ and $\mathcal{C}(\mathcal{V}_d)$.
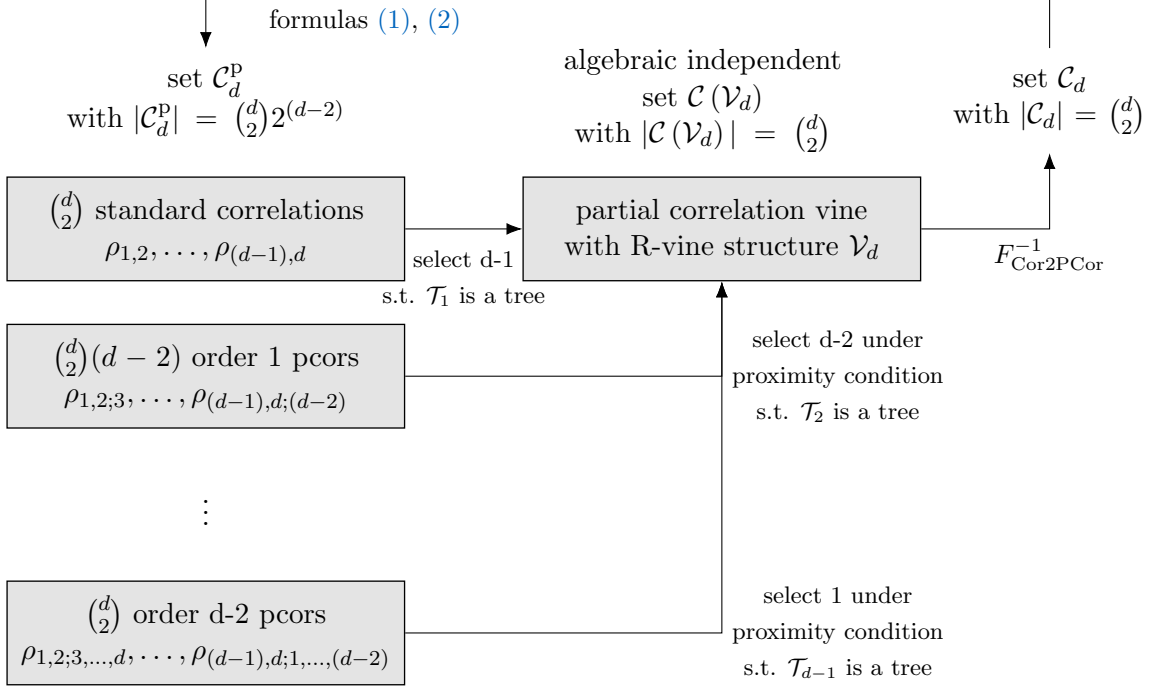
Figure 3: Illustration of the transformation of the set of standard correlations $\mathcal{C}_d$ through a partial correlation vine, which consists of a subset of algebraic independent (partial) correlations $\mathcal{C}(\mathcal{V}_d) \subset \mathcal{C}_d^{\mathrm{p}}$ from all standard and partial correlations. The abbreviation "pcor" is used for partial correlation.

3 General setting and benchmark models

In the following, partial correlation vine based data transformation will be used to model and forecast multivariate volatility time-series. To do so, we introduce the general data setting first. For the daily price series $\boldsymbol{S}_t \in \mathbb{R}^d$, $t = 1, \ldots, T$, of d assets let $\boldsymbol{r}_t = \log(\boldsymbol{S}_t) - \log(\boldsymbol{S}_{t-1})$ be the $d \times 1$ vector of daily log-returns. The process $\boldsymbol{r}_t$ can be written as

$$\boldsymbol{r}_t = \mathbb{E}[\boldsymbol{r}_t | \mathcal{F}_{t-1}] + \boldsymbol{\epsilon}_t,$$

where $\mathcal{F}_{t-1}$ is the information set containing all information up to and including time point $t-1$. For the innovation term $\boldsymbol{\epsilon}_t$, we suppose that $\boldsymbol{\epsilon}_t = \boldsymbol{\Sigma}_t^{1/2} \boldsymbol{\eta}_t$, where $\boldsymbol{\Sigma}_t = \mathrm{Var}[\boldsymbol{r}_t | \mathcal{F}_{t-1}]$ is the $(d \times d)$-dimensional symmetric and positive definite conditional covariance matrix. For the i.i.d. vector $\boldsymbol{\eta}_t \in \mathbb{R}^d$ it holds that $\mathbb{E}[\boldsymbol{\eta}_t] = 0$ and $\mathrm{Var}[\boldsymbol{\eta}_t] = I_d$. Interest is in modeling and forecasting the series of daily conditional covariance matrices $\boldsymbol{\Sigma}_t$, $t = 1, \ldots, T$, which however are naturally latent variables and therefore are unobservable. Still, as proposed by Barndorff-Nielsen and Shephard (2004) $\boldsymbol{\Sigma}_t$, $t = 1, \ldots, T$, can be specified nonparametrically using the realized covariance matrices as consistent estimates. Considering M intra-day periods per day t, the latter are calculated from high-frequency intra-day log-returns $\boldsymbol{r}_{\ell,t} = \log(\boldsymbol{S}_{t-1+\ell/M}) - \log(\boldsymbol{S}_{t-1+(\ell-1)/M})$ based on the price series $\boldsymbol{S}_{\ell,t} \in \mathbb{R}^d$, $\ell = 1, \ldots, M$. The modeling and forecasting framework is then based on the matrix valued time-series of realized covariance matrices

$$\boldsymbol{Y}_t = \sum_{l=1}^{M} \boldsymbol{r}_{\ell,t} \boldsymbol{r}_{\ell,t}', \quad t = 1, \ldots, T. \tag{5}$$

Since for the matrix forecasts symmetry and positive definiteness have to be ensured, algebraic restrictions are imposed on time-series models. Thus, popular modeling strategies avoid direct modeling of the realized covariance matrices considering transformed data instead. Then, the modeling approach basically consists of three consecutive steps: (S1) data transformation of the realized covariance matrices; (S2) multivariate time-series modeling and prediction based on the transformed data; (S3) back-transformation of the transformed data to obtain predictions for the realized covariance matrices, which are proxies for the future conditional covariance matrices.

The novelty in this paper lies in the use of partial correlation vines for data transformation in steps (S1) and (S3). By modeling and forecasting the time-series of partial correlation vines, we obtain forecasts for the transformed data, which do not underly any algebraic restrictions. On the contrary, due to the algebraic independence of the model components positive definiteness of the corresponding predicted correlation matrices is always guaranteed. In literature, besides the matrix log transformation suggested by Bauer and Vorkink (2011) data transformation based on the Cholesky factorization is one of the most commonly used approaches.

Here, the series of realized covariance matrices $\boldsymbol{Y}_t$, $t = 1, \ldots, T$, is decomposed such that $\boldsymbol{Y}_t = \boldsymbol{C}_t' \boldsymbol{C}_t$, where $\boldsymbol{C}_t'$ is a lower triangular matrix with positive diagonal elements. The Cholesky elements $c_{i,j;t}$ $(i, j = 1, \ldots, d)$ are recursively calculated by

$$
c_{i,j;t} = \begin{cases} \frac{1}{c_{i,i;t}} \left(y_{i,j;t} - \sum_{k=1}^{i-1} c_{k,i;t} c_{k,j;t} \right) & \text{for } i < j, \\ \sqrt{y_{j,j;t} - \sum_{k=1}^{j-1} c_{k,j;t}^2} & \text{for } i = j, \\ 0 & \text{for } i > j. \end{cases}
\tag{6}
$$

By modeling and forecasting the Cholesky elements in step (S2) no parameter restrictions need to be imposed on the multivariate time-series models. Symmetry and positive definiteness of the predicted covariance matrices $\hat{\boldsymbol{Y}}_t$, $t = T + 1, T + 2, \ldots$, are automatically guaranteed through the back-transformation

$$
\hat{y}_{i,j;t} = \sum_{k=1}^{\min\{i,j\}} \hat{c}_{k,i;t} \hat{c}_{k,j;t}.
\tag{7}
$$

Chiriac and Voev (2011) use the Cholesky decomposition in steps (S1) and (S3) and apply a parsimonious VARFIMA model to the multivariate time-series of the Cholesky components in step (S2). In their detailed analysis, they show the superiority of their approach over a variety of competitor models. The comparison includes the above mentioned matrix log transformation used in steps (S1) and (S3) for data transformation combined with VARFIMA and vector HAR models in step (S2). Further, the Wishart autoregressive model of Gouriéroux et al. (2009) as well as the multivariate GARCH model with dynamic conditional correlations of Engle (2002) and its fractionally integrated version proposed by Baillie et al. (1996) are considered. Brechmann et al. (2018) refine the Cholesky-VARFIMA model of Chiriac and Voev (2011) and allow for more flexible modeling of the multivariate time-series in step (S2). They take account of challenging data characteristics in the Cholesky elements by modeling the univariate marginal time-series with elaborate HAR and ARFIMA models including GARCH-augmentations for the residuals. The possibly complex dependence between the Cholesky components is captured by a copula. Given these profound model reviews and comparisons already existing in literature, models based on the Cholesky decomposition will be our main benchmarks.

Chiriac and Voev (2011) and Brechmann et al. (2018) both consider high frequency data from the NYSE TAQ database containing tick-by-tick bid and ask quotes on six stocks listed

on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX) and the National Association of Security Dealers Automated Quotation System (NASDAQ). The original raw data were processed by Chiriac and Voev (2011), who provide detailed information on the employed data preparation. Data of the six stocks American Express Inc. (AXP), Citigroup (C), General Electric (GE), Home Depot Inc. (HD), International Business Machines (IBM) and JPMorgan Chase & Co (JPM) were sampled from 9:30 until 16:00 for the period January 1, 2000, until July 30, 2008, i.e. for 2156 trading days. While in (5) a single realized covariance matrix is computed from M intra-day log-returns, Chiriac and Voev (2011) obtained for each day a refined subsampled realized covariance matrix, which is more robust to market microstructure noise. For each day t, a 5-minute spaced time grid, i.e. $M = 78$, was shifted by 10 seconds, resulting in 30 distinct sets of realized covariance matrices calculated from 78 intra-day log-returns. By taking the average of these sets, the subsampled realized covariance matrix for day t was calculated. Although the data are less recent, we consider the same data for comparison reasons. Further, the data cover interesting periods of financial turmoil such as the aftermath of the dotcom bubble in 2000 and the beginning of the financial crisis in 2008. Since focus in this paper is on the novel data transformation defined by partial correlation vines, this will provide new interesting insights about the data.

4 Partial correlation vine data transformation approach

In this section, we outline – supported by real data characteristics – steps (S1) to (S3) for the proposed modeling strategy based on partial correlation vines.

4.1 Data characteristics

Time-series of realized co(variances) typically exhibit long-memory behavior detectable by high autocorrelations, which decay at a slow rate (see e.g. Andersen and Bollerslev (1997); Andersen et al. (2001)). Chiriac and Voev (2011) find that the time-series of Cholesky components obtained through data transformation inherit this data feature. Further, according to Brechmann et al. (2018) appropriate time-series models need to capture non-Gaussianity and volatility clustering of the residuals extracted from the series of Cholesky elements.

In order to also appropriately setup the partial correlation vine data transformation model it is essential to understand the properties of the corresponding model components, namely realized variances and realized (partial) correlations. The latter are specified through the realized covariance matrix via $\boldsymbol{Y}_t = \boldsymbol{D}_t^{1/2} \boldsymbol{R}_t \boldsymbol{D}_t^{1/2}$, $t = 1, \ldots, T$. For day t, $\boldsymbol{D}_t = \text{diag}(y_{1,1;t}, \ldots, y_{d,d;t})$ contains the realized variances and $\boldsymbol{R}_t$ is the realized correlation matrix. Realized partial correlations can easily be obtained either using recursion formula (1) or through simultaneous calculation using (2). For reasonable time-series modeling later in step (S2), for all model components data on the real line are needed. Thus, we log-transform the all positive realized variance time-series and apply the Fisher z-transformation to the series of (partial) correlations, i.e. for ρ_t being an arbitrary (partial) correlation at day t

$$z(\rho_t) = \frac{1}{2} \log\left(\frac{1 + \rho_t}{1 - \rho_t}\right), \quad t = 1, \ldots, T. \tag{8}$$

For the considered real data, Figure 4 shows a selection of time-series both on the original (left) and the transformed (right) scale. The first panel illustrates for JPM the daily realized variance series. Striking is the highly volatile behavior particularly during periods of financial turmoil such as the aftermath of the dotcom bubble and the beginning of the financial crisis in August 2007. Panels 2 to 6 show selected daily time-series of realized (partial) correlations with increasing order, e.g. the last panel illustrates the time-series of the fourth order realized partial correlation between HD and JPM. For each day, the latter is a proxy of the conditional
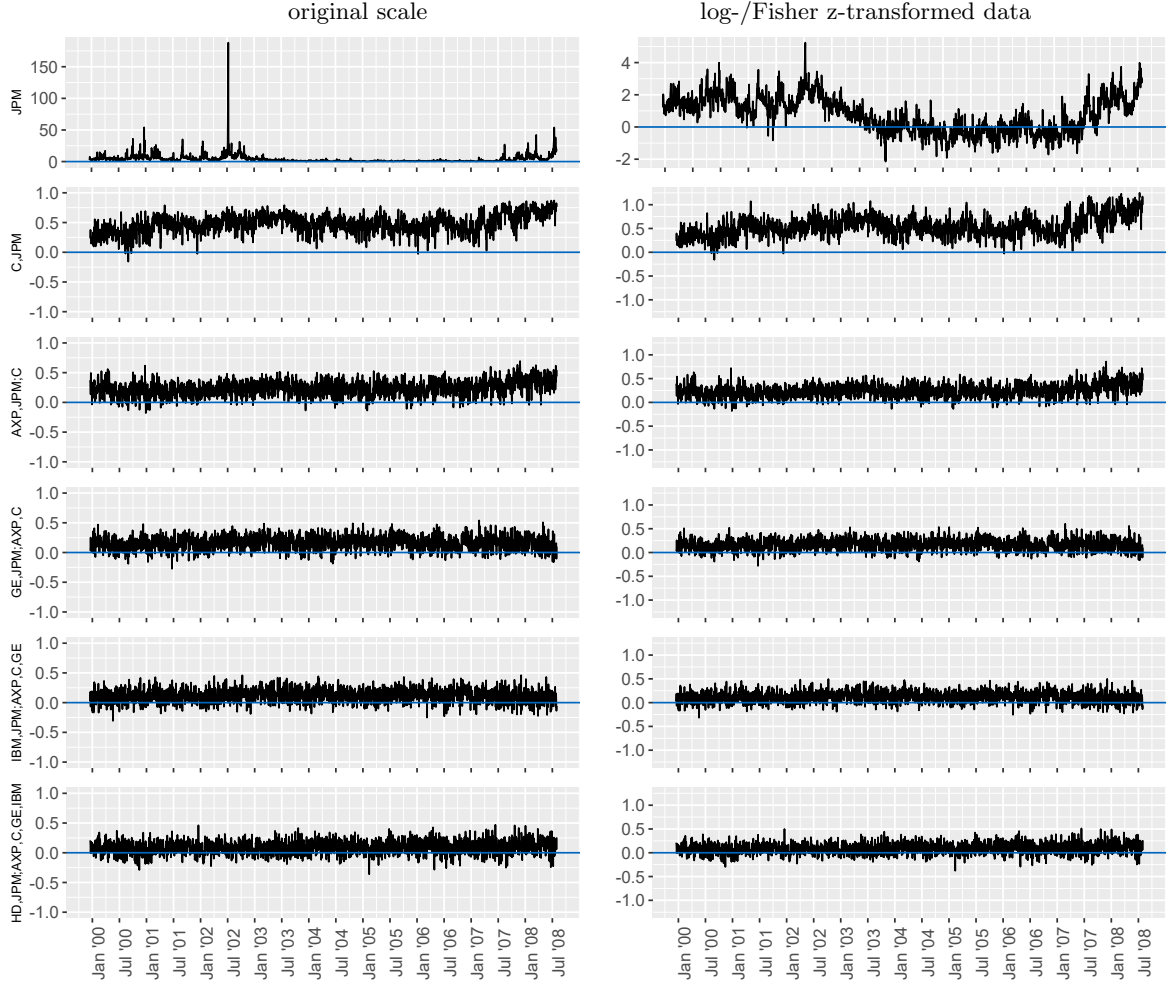
Figure 4: Daily realized variance series (1st row) and daily realized (partial) correlation series (2nd – 6th row). Original data are shown in the left panel, log-transformed and Fisher z-transformed data, respectively, are shown in the right panel.

(with respect to the information set) correlation between the log-returns of HD and JPM given the four remaining stocks. With increasing order the realized partial correlation time-series become more stable while still exhibiting highly volatile behavior. Further, in Figure 5 data characteristics of four time-series are illustrated. The figures in the top row correspond to the realized variance time-series of JPM, which together with the remaining five realized variance series always will be a model component. The long hyperbolic decay of the autocorrelation function of the squared data on the left confirms the long-memory behavior and the presence of volatility clustering. The log-periodogram shows higher peaks only for short frequencies as expected for self-similar processes. In the second and third row, exemplary time-series, which would appear in tree level $\mathcal{T}_1$ and $\mathcal{T}_4$, respectively, of an R-vine structure are shown. Interestingly, while the realized standard correlation time-series corresponding to tree level $\mathcal{T}_1$ inherits the data characteristics of the realized variance time-series, the latter are less pronounced for the realized third order partial correlation time-series in $\mathcal{T}_4$.

To gain a deeper understanding of this last observation, recall that in a partial correlation vine each variable pair (i, j) $(i, j \in \{1, \ldots, d\}, i \neq j)$ forms exactly once the conditioned set of an edge. Thus, depending on the tree level ℓ the proxy for the conditional (with respect to the information set) correlation between the log-returns of stocks i and j is either represented by the realized standard correlation (if (i, j) occurs as conditioned set in $\mathcal{T}_1$) or through a $(\ell-1)$-th order realized partial correlation (if (i, j) occurs as conditioned set in $\mathcal{T}_\ell$ $(\ell = 2, \ldots, d-1)$). In the latter case, the linear effect of the $\ell-1$ stocks forming the conditioning set is removed.
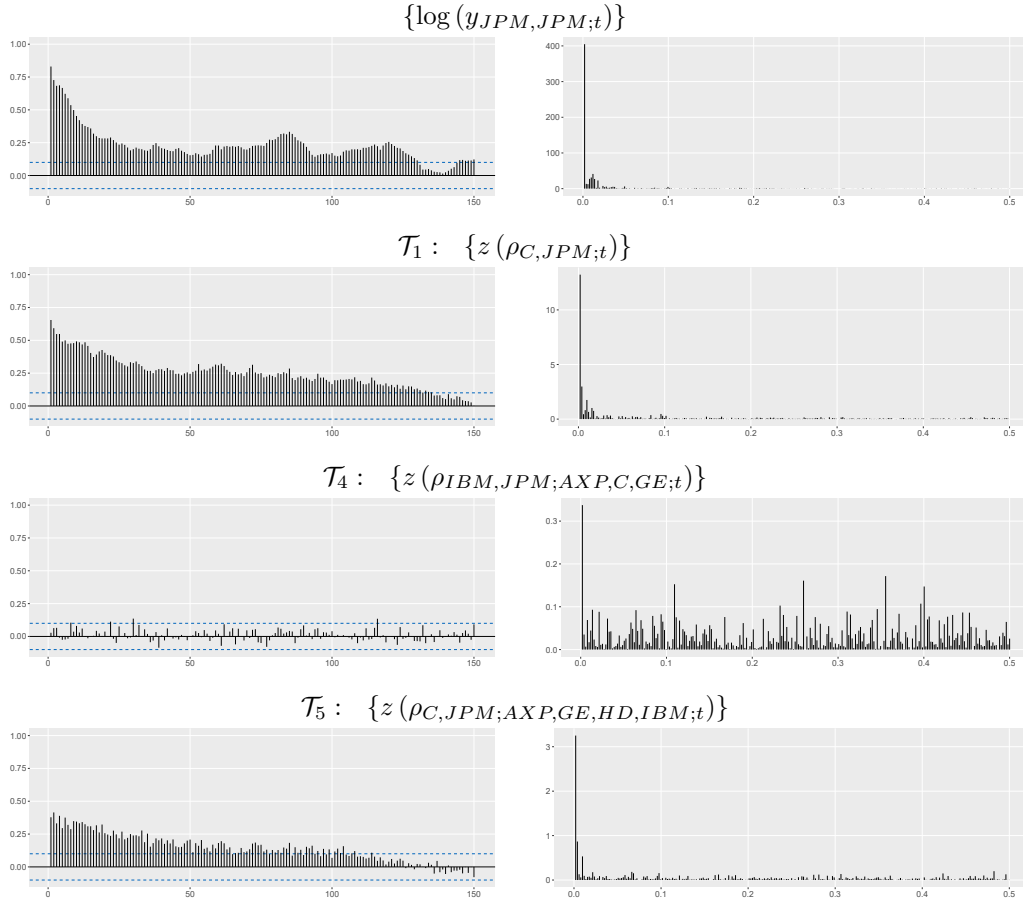
Figure 5: Illustration of autocorrelation functions of squared data (left panel) and corresponding log-periodograms (right panel) based on data from July 1, 2006, to June 30, 2008. In the first row, the log-transformed realized variance time-series of JPM is considered. In rows 2 to 4 exemplary Fisher z-transformed realized (partial) correlation time-series of increasing order are considered.

Clearly, for some pairs the realized standard correlations might mainly be driven by other variables. Once this influence is removed data features such as long-memory behavior weaken and the corresponding realized partial correlation time-series behave more and more like noise. On the other hand, this effect is not observable for pairs, which truly are strongly correlated such as the log-returns of the two financial stocks C and JPM. The corresponding realized fourth order partial correlation time-series, which would occur in the highest tree level, i.e. $\mathcal{T}_5$, of an R-vine structure, underlies the two figures in the last row of Figure 5. It shows similar data characteristics as the realized variance time-series in the top row.

The above analysis clearly stresses the practical interpretability of the model components in the partial correlation vine data transformation approach, namely realized variances and realized (partial) correlations. In the following, the detected inhomogeneous data complexity motivates a specific choice for the R-vine structure used for data transformation in step (S1).

4.2 Step (S1): R-vine structure selection for data transformation

In d dimensions there exist $d!/2 \cdot d^{(d-2)(d-3)/2}$ valid R-vines (Morales Napoles et al. 2010). It is important to note that, in general, each of these tree structures allows a valid transformation of the realized correlation matrices. To allow for model parsimony later in step (S2), when modeling the dynamics of the realized variance series and the realized (partial) correlation time-series selected by the R-vine structure, we refer to the data characteristics detected in Section 4.1 and propose in this section an algorithm for R-vine structure selection.

We know that when transforming a series of realized correlation matrices based on the

same R-vine structure, to each edge in this R-vine a univariate time-series of realized standard or partial correlations is assigned. Therefore, each edge can be characterized by a weight derived from sample properties of the corresponding time-series. We decide for the average (partial) correlation strength: we consider the average correlation matrix $\bar{\boldsymbol{R}} = (\bar{\rho}_{i,j})_{i,j=1,\ldots,d}$ (which is positive definite) calculated from $\boldsymbol{R}_t$, $t = 1, \ldots, T$. Then, we find the maximum spanning tree $\mathcal{T}_1$ (Katoh et al. 1981) with edge weights set to $\bar{\rho}_{C_{e,a},C_{e,b}}$. To construct tree $\mathcal{T}_2$, we calculate all average first order partial correlations $\bar{\rho}_{C_{e,a},C_{e,b};D_e}$, i.e. $|D_e| = 1$, where $(C_{e,a}, C_{e,b}; D_e)$ satisfies the proximity condition given $\mathcal{T}_1$. Based on these weights, we find the maximum spanning tree $\mathcal{T}_2$. In general, we construct the R-vine structure $\mathcal{V}_d$ within a top-down procedure and find tree by tree ($\ell = 1, \ldots, d-1$) the maximum spanning tree $\mathcal{T}_\ell$ with edge weights set to $\bar{\rho}_{C_{e,a},C_{e,b};D_e}$, where $|D_e| = \ell - 1$ and $(C_{e,a}, C_{e,b}; D_e)$ satisfies the proximity condition given $\mathcal{T}_1$ to $\mathcal{T}_{\ell-1}$. By doing so, we equip based on historical information the R-vine structure with the highest realized (partial) correlation means.

The correlation matrix $\bar{\boldsymbol{R}}$ can be obtained in various ways depending on how the average is calculated. Considering for each pair (i, j) ($i, j \in \{1, \ldots, d\}$, $i \neq j$) the empirical mean $\bar{\rho}_{i,j} = \frac{1}{T} \sum_{t=1}^{T} \rho_{i,j;t}$ assigns to each day's value $\rho_{i,j;t}$ the same influence $1/T$ irrespective of how far it lies in the past. By using e.g. an exponentially weighted moving average (EWMA) more influence can be assigned to values of more recent days. Here, the exact weights are controlled by the smoothing parameter $\lambda \in]0, 1[$ and are defined as $w_t = (1 - \lambda)\lambda^{T-t}$, $t = 1, \ldots, T$. Thus, for decreasing λ the impact of more recent days increases and therewith the sensitivity of the selected R-vine structure with respect to market changes.

For the real data example, the proposed R-vine structure selection method is illustrated in Table 1. Recall that the data include three market participants of financial sectors, namely AXP, C and JPM, IBM as an IT service, HD representing building materials trade and the diversified industrial corporation GE. As edge weights the empirical means of the realized (partial) correlation series based on all data points, i.e. $t = 1, \ldots, 2156$ (January 1, 2000 - July 30, 2008), are chosen. In $\mathcal{T}_1$, we start with a full graph, i.e. all edges are allowed to be chosen. Edge by edge a tree, i.e. a connected and acyclic graph, is built adding edges with the highest possible correlation mean. Including e.g. the pair (AXP,JPM) would result in a cycle and is thus not allowed. For $\mathcal{T}_2$ only edges satisfying the proximity condition given $\mathcal{T}_1$ are allowed. A tree is constructed by the four edges with the highest mean values of realized first order partial correlations, etc. The resulting R-vine structure captures the strong pairwise realized correlations between the three financial services in trees $\mathcal{T}_1$ and $\mathcal{T}_2$. From $\mathcal{T}_3$ on only realized partial correlations corresponding to stocks from different market sectors are modeled such as the one investigated in the last row of Figure 5. Consequently, using in step (S1) an R-vine structure selected by the proposed algorithm will likely result in higher order realized partial correlation series, which allow for a parsimonious time-series model specification in step (S2).

4.3 Step (S2): Multivariate time-series modeling and forecasting

After transforming the series of realized covariance matrices in step (S1), multivariate time-series models in step (S2) can be applied to the transformed data without imposing any parameter restrictions. Except for the considered modeling components this step does not differ from Cholesky decomposition based benchmark models. In both approaches, there are $d(d+1)/2$ model components after data transformation. In particular, the time-series of log-transformed realized variances and Fisher z-transformed realized (partial) correlations could be modeled using a VARFIMA model as suggested for the Cholesky elements in Chiriac and Voev (2011). Compared with this, copula based time-series modeling as applied in Brechmann et al. (2018) showed superior results especially for economic applications.

A $\tilde{d}$-dimensional copula is a multivariate distribution function on $[0, 1]^{\tilde{d}}$ with uniformly distributed margins. Since data are required to be approximately i.i.d., the copula model usually is not directly applied to the observed time-series, but to the corresponding standardized

Table 1: Illustration of the R-vine structure selection method for the real data example considering all available data points, i.e. the mean values $\bar{\rho}_{\mathcal{C}_{e,a},\mathcal{C}_{e,b};\mathcal{D}_e}$ are based on $t = 1, \ldots, 2156$.

pairs allowed by proximity condition			selected tree
$\mathcal{D}_e$	$\mathcal{C}_{e,a},\mathcal{C}_{e,b}$	$\bar{\rho}_{\mathcal{C}_{e,a},\mathcal{C}_{e,b};\mathcal{D}_e}$	
$\emptyset$	C,JPM	0.547	
$\emptyset$	AXP,C	0.456	
$\emptyset$	C,GE	0.437	
$\emptyset$	AXP,JPM	0.433	
$\emptyset$	GE,IBM	0.400	
$\emptyset$	AXP,GE	0.394	
$\emptyset$	GE,JPM	0.393	
$\emptyset$	C,IBM	0.390	$\mathcal{T}_1$
$\emptyset$	IBM,JPM	0.362	
$\emptyset$	AXP,IBM	0.358	
$\emptyset$	C,HD	0.355	
$\emptyset$	GE,HD	0.352	
$\emptyset$	AXP,HD	0.333	
$\emptyset$	HD,JPM	0.333	
$\emptyset$	HD,IBM	0.330	
GE	C,IBM	0.253	
C	AXP,JPM	0.247	
C	AXP,GE	0.241	
C	GE,HD	0.229	$\mathcal{T}_2$
C	GE,JPM	0.214	
C	AXP,HD	0.203	
C	HD,JPM	0.182	
C,GE	HD,IBM	0.164	
C,GE	AXP,IBM	0.163	
AXP,C	GE,JPM	0.163	$\mathcal{T}_3$
C,GE	AXP,HD	0.154	
C,GE,IBM	AXP,HD	0.129	
AXP,C,GE	IBM,JPM	0.121	$\mathcal{T}_4$
AXP,C,GE,IBM	HD,JPM	0.093	$\mathcal{T}_5$

15

residuals $\left(\varepsilon_{1;t}, \ldots, \varepsilon_{\tilde{d};t}\right)$, $t = 1, \ldots, T$. The latter are extracted after fitting appropriate univariate time-series models to the original marginal data. While no longer being subject to temporal dependence, the residuals inherit the cross-sectional dependence between the time-series components. According to Sklar (1959), their joint distribution function F can be expressed in terms of its marginal distributions F_j ($j = 1, \ldots, \tilde{d}$) and its corresponding copula, i.e. $F\left(\varepsilon_1, \ldots, \varepsilon_{\tilde{d}}\right) = \mathbb{C}\{F_1\left(\varepsilon_1\right), \ldots, F_{\tilde{d}}\left(\varepsilon_{\tilde{d}}\right)\}$. Consequently, in a copula based time-series model the individual behavior of the time-series components and their dependence are modeled separately. This allows us to deepen the analysis of the realized variance and (partial) correlation series.

Marginal time-series modeling

As discussed in Section 4.1 specific univariate time-series models are needed to reproduce the long-memory property of the Cholesky components as well as of the realized variance and some of the realized (partial) correlation series. HAR (Corsi 2009) and ARFIMA (Andersen et al. 2003) models are popular models capable of doing so.

Let η_t denote the variable of interest, i.e. a log-transformed realized variance, a Fisher z-transformed realized (partial) correlation or a Cholesky element, at time t. A basic HAR model accounts for different time horizons by incorporating one day ($d = 1$), one week ($w = 5$) and one month ($m = 22$) averages η_{t-1}, $\eta_{t-1}^{(w)}$ and $\eta_{t-1}^{(m)}$ as regressors for η_t:

$$\eta_t = \alpha_0 + \alpha_1 \eta_{t-1} + \alpha_2 \eta_{t-1}^{(5)} + \alpha_3 \eta_{t-1}^{(22)} + \epsilon_t.$$

The error term ϵ_t is usually assumed to be Gaussian white noise. While showing very good modeling and prediction performance given complex data features, the basic HAR model describes an easy to estimate restricted autoregressive process.

The ARFIMA(p, D, q) model for the time-series η_t, $t = 1, \ldots, T$, is specified by

$$\phi\left(L\right)\left(1 - L\right)^D \left(\eta_t - \mu\right) = \psi\left(L\right)\epsilon_t,$$

where $\phi\left(L\right) = 1 - \phi_1 L - \ldots - \phi_p L^p$ and $\psi\left(L\right) = 1 + \psi_1 L + \ldots + \psi_q L^q$ are lag polynomials for $p, q \in \mathbb{N}$. D is the parameter of fractional differencing. We choose $D \in (0, 0.5)$ to guarantee stationarity of the process. Gaussian white noise is usually assumed for the error term ϵ_t.

In these basic models, the volatility h of the error term $\epsilon_t = h\varepsilon_t$ with $\varepsilon_t \sim \mathcal{N}\left(0, 1\right)$ is assumed to be constant. Given the presence of volatility clustering in the Cholesky series, Brechmann et al. (2018) include a GARCH$\left(1, 1\right)$ component, i.e. $\epsilon_t = h_t \varepsilon_t$ with $h_t^2 = \omega + \beta_1 \epsilon_{t-1}^2 + \beta_2 h_{t-1}^2$. Usually, the innovation terms ε_t are standard normally distributed, i.e. $\varepsilon_t \sim \mathcal{N}\left(0, 1\right)$. To additionally capture possible high kurtosis and skewness, Brechmann et al. (2018) further allow the innovations to follow a skewed generalized error distribution, i.e. $\varepsilon_t \sim \text{SGED}\left(\mu, \sigma, \nu, \xi\right)$ (Bai et al. 2003; Corsi et al. 2008; Fernández and Steel 1998). A specification of the skewed generalized error distribution is provided in Appendix A.

Dependence modeling

After fitting one of the above univariate time-series models to each of the model components, the sample of i.i.d. standardized residuals $\left(\varepsilon_{1;t}, \ldots, \varepsilon_{d(d+1)/2;t}\right)$, $t = 1, \ldots, T$, can be extracted. To them, usually a two-stage proceeding is applied, called inference for margins methods (Joe and Xu 1996; Joe 2005). First, the probability integral transform, $\hat{u}_{j;t} = \hat{F}_j\left(\varepsilon_{j;t}\right)$, is applied to each residual component ($j = 1, \ldots, d(d + 1)/2$) to obtain pseudo copula data $\left(\hat{u}_{1;t}, \ldots, \hat{u}_{d(d+1)/2;t}\right)$, $t = 1, \ldots, T$. The marginal estimates $\hat{F}_j$ are specified through the corresponding marginal time-series fit. In case of e.g. a basic HAR or ARFIMA model, $\hat{F}_j$ is a normal distribution with sample mean (approximately 0) and sample standard deviation (approximately 1).

Second, a copula is fitted to the pseudo copula data. To do so, we consider R-vine copulas. Note that while in step (S1) and (S3) of the partial correlation vine data transformation approach, R-vines are exclusively used as a graph theoretical tool for data transformation, they are now the cornerstones for this flexible copula class. Given the limited choice of copula families for more than two variables, Joe (1996) proposed to recursively decompose the multivariate copula density into a cascade of (conditional) bivariate copula densities through conditioning. Through R-vines the variety of possible pair-copula constructions can be organized. To each edge in an R-vine structure bivariate unconditional copulas in tree $\mathcal{T}_1$ and bivariate conditional copulas in trees $\mathcal{T}_2$ to $\mathcal{T}_{d-1}$ with arbitrary type and strength of dependence are assigned. This is allows to capture complex asymmetric and nonlinear dependence patterns. A detailed introduction to R-vine copulas is provided in Appendix B. Fitting an R-vine copula to the sample $\left(\hat{u}_{1;t}, \ldots, \hat{u}_{d(d+1)/2;t}\right)$, $t = 1, \ldots, T$, finalizes the model specification based on in-sample data.

Forecasting of the model components

A one-day-ahead out-of-sample forecast $\hat{\boldsymbol{Y}}_{T+1}$ is now generated in multiple steps. First, innovations on the copula scale $\left(\hat{u}_{1;T+1}, \ldots, \hat{u}_{d(d+1)/2;T+1}\right)$ are sampled from the R-vine copula fit. The corresponding innovations on their original scale are obtained using the inverse probability integral transform, i.e. $\hat{\varepsilon}_{j;T+1} = \hat{F}_j^{-1}\left(\hat{u}_{j;T+1}\right)$ $(j = 1, \ldots, d(d+1)/2)$. Then, based on the corresponding time-series fit forecasts for the model components, which involve the corresponding simulated innovations, are calculated. In the Cholesky decomposition based model, this results in a predicted upper triangular matrix $\hat{\boldsymbol{C}}_{T+1}$. In the partial correlation vine data transformation approach, the log-transformation of the realized variances and the Fisher z-transformation of the realized (partial) correlations have to be reversed first. This results in the predicted realized partial correlation vine stored in $\hat{\boldsymbol{P}}_{\mathcal{C}(\mathcal{V}_d);T+1}$ and the corresponding predicted realized variance vector $(\hat{y}_{1,1;T+1}, \ldots, \hat{y}_{d,d;T+1})$.

4.4 Step (S3): Back-transformation

Finally, based on $\hat{\boldsymbol{C}}_{T+1}$ back-transformation (7) is applied for the Cholesky decomposition based approach. Likewise, $\hat{\boldsymbol{P}}_{\mathcal{C}(\mathcal{V}_d);T+1}$ is back-transformed to a symmetric and positive definite correlation matrix (Section 2) based on the R-vine structure selected in step (S1) (Section 4.2). Combined with the predicted realized variances a forecast for the realized covariance matrix is obtained. Given that both backward procedures involve nonlinear transformations of the copula-distributed innovation terms, the underlying dependence pattern has an explicit effect on the matrix forecast. Clearly, in practice this simulation based procedure is to be replicated several times. The final point-forecast $\hat{\boldsymbol{Y}}_{T+1}$, which is considered as a proxy for the conditional covariance matrix, is obtained as the mean of the simulation based matrix forecasts.

In both modeling approaches the predictions of $\hat{\boldsymbol{Y}}_{T+1}$ are obtained after inverting a nonlinear data transformation. Consequently, while the prediction errors of the model components have zero mean, the nonlinear back-transformation induces a bias. Even though Chiriac and Voev (2011) derive the theoretical bias correction for the Cholesky decomposition based model, they stress that the theoretical formula crucially depends on the considered time-series model and thus, has to be estimated in practice. However, given that in a copula based time-series model the marginal time-series are estimated independently of each other, consistent estimation of the covariance matrix of the forecast errors in Brechmann et al. (2018) is not feasible. Against this background, Chiriac and Voev (2011) and Brechmann et al. (2018) both advocate a data-driven bias correction. In the partial correlation vine data transformation approach, the forecast bias of the variable pair (i, j) in $\hat{\boldsymbol{Y}}_{T+1}$ depends not only on the underlying time-series model but also on the R-vine structure used for data transformation, making a theoretical correction practically infeasible. We therefore as well opt for

the heuristic data-driven bias correction proposed in Chiriac and Voev (2011). The basic idea is to match the level of the observed volatilities by scaling the predicted volatilities $\sqrt{\hat{y}_{j,j;T+1}}$ ($j = 1, \ldots, d$) by the corresponding mean $\frac{1}{T-s+1} \sum_{t=s}^{T} \frac{\sqrt{y_{j,j;t}}}{\sqrt{\hat{y}_{j,j;t}}}$, where s controls the number of past days included for level matching. Note that this proceeding has no influence on the predicted correlation structure. Thus, in the partial correlation vine data transformation approach only the nonlinear inversion of the log-transformation can be corrected.

4.5 Modeling approach at a glance

Figure 6 summarizes the partial correlation vine data transformation approach discussed in the previous sections.

5 Empirical study

The real data example introduced in Section 3 and Section 4 will now be investigated in more detail. Based on the model specifications in Section 5.2 and Section 5.3, the out-of-sample forecasting performance of the partial correlation vine data transformation approach and Cholesky decomposition based benchmark models will be evaluated both with respect to statistical precision and mean-variance trade-off in portfolio optimization strategies.

It is crucial to keep in mind that the realized covariance matrices are proxies for the unobservable true conditional covariance matrices, which we aim to predict. As a consequence, when comparing the performance of different forecasting models, loss functions have to satisfy the condition to deliver the same ranking whether the evaluation is based on the unbiased proxy, i.e. the realized covariance matrix, or the true conditional covariance matrix. We will therefore rely on loss functions, which according to Patton (2011) and Laurent et al. (2013) are robust to noise in the volatility proxies. Further, numerous different models for prediction will be compared. To avoid pairwise comparison of loss functions we apply the model confidence set (MCS) approach developed by Hansen et al. (2011). It allows to select a set of superior models, which contains the best one with a specified level of confidence.

5.1 Moving window approach

In the following, we proceed in a moving window approach. Data for the period from January 1, 2000, until June 30, 2008, are available, i.e. for 2156 days. For each time window 502 days (about two years) are used as training set and 22 days (about one month) constitute the test set for which one-day-ahead forecasts are made. Since in case of HAR based time-series models a monthly (22 days) average of the data is involved, the first forecast is obtained for day 525. In total, there are 75 time windows. Figure 7 illustrates the moving window approach.

5.2 Step (S1): Dynamic data transformation

For each time window W_i ($i = 1, \ldots, 75$) the realized covariance matrices of the corresponding training set are transformed in step (S1). Clearly, application of the R-vine structure selection algorithm proposed in Section 4.2 can lead to varying R-vine structures among time windows. Thus, data transformation in the partial correlation vine data transformation approach may dynamically change over time. Depending on how the average correlation matrix used for R-vine structure selection is calculated, the selected R-vine structure is more or less sensitive to market developments.

Data: series of daily realized covariance matrices $\boldsymbol{Y}_t$, $t = 1, \ldots, T$

Figure 1

Figure 3, Section 4.2

Series of realized variance vectors $\boldsymbol{y}_t$, $t = 1, \ldots, T$

Series of realized (partial) correlation vectors specified by $\mathcal{V}_d$ $\boldsymbol{P}_{\mathcal{C}(\mathcal{V}_d);t}$, $t = 1, \ldots, T$

log-transformation

Fisher z-transformation

Series of model components scaled to $\mathbb{R}^{d(d+1)/2}$

Section 4.3

Univariate time-series modeling resulting in an approximately i.i.d. sample of innovations $(\varepsilon_{1;t}, \ldots, \varepsilon_{d(d+1)/2;t})$, $t = 1, \ldots, T$, and pseudo copula data $(\hat{u}_{1;t}, \ldots, \hat{u}_{d(d+1)/2;t})$, $t = 1, \ldots, T$

Appendix B

Fit an R-vine copula to $(\hat{u}_{1;t}, \ldots, \hat{u}_{d(d+1)/2;t})$, $t = 1, \ldots, T$

set $i = 1$

Simulate copula data $(\hat{u}_{1;T+1}, \ldots, \hat{u}_{d(d+1)/2;T+1})$ according to the R-vine copula fit and calculate simulated innovations $(\hat{\varepsilon}_{1;T+1}, \ldots, \hat{\varepsilon}_{d(d+1)/2;T+1})$

Calculate predicted model components scaled to $\mathbb{R}^{d(d+1)/2}$

exp-transformation

inverse Fisher z-transformation

Predicted realized variance vector $(\hat{y}_{1,1;T+1}, \ldots, \hat{y}_{d,d;T+1})$

Predicted realized partial correlation vine stored in $\hat{\boldsymbol{P}}_{\mathcal{C}(\mathcal{V}_d);T+1}$

Section 2, F_{Cor2PCor}^{-1}

Predicted realized correlation matrix $\hat{\boldsymbol{R}}$

Predicted realized covariance matrix based on i-th simulated innovation vector $\hat{\boldsymbol{Y}}_{T+1}^i$

$i = i + 1$

no

$i = N$

yes

One-day-ahead forecast $\hat{\boldsymbol{Y}}_{T+1}$ obtained as mean of $\hat{\boldsymbol{Y}}_{T+1}^i$, $i = 1, \ldots, N$
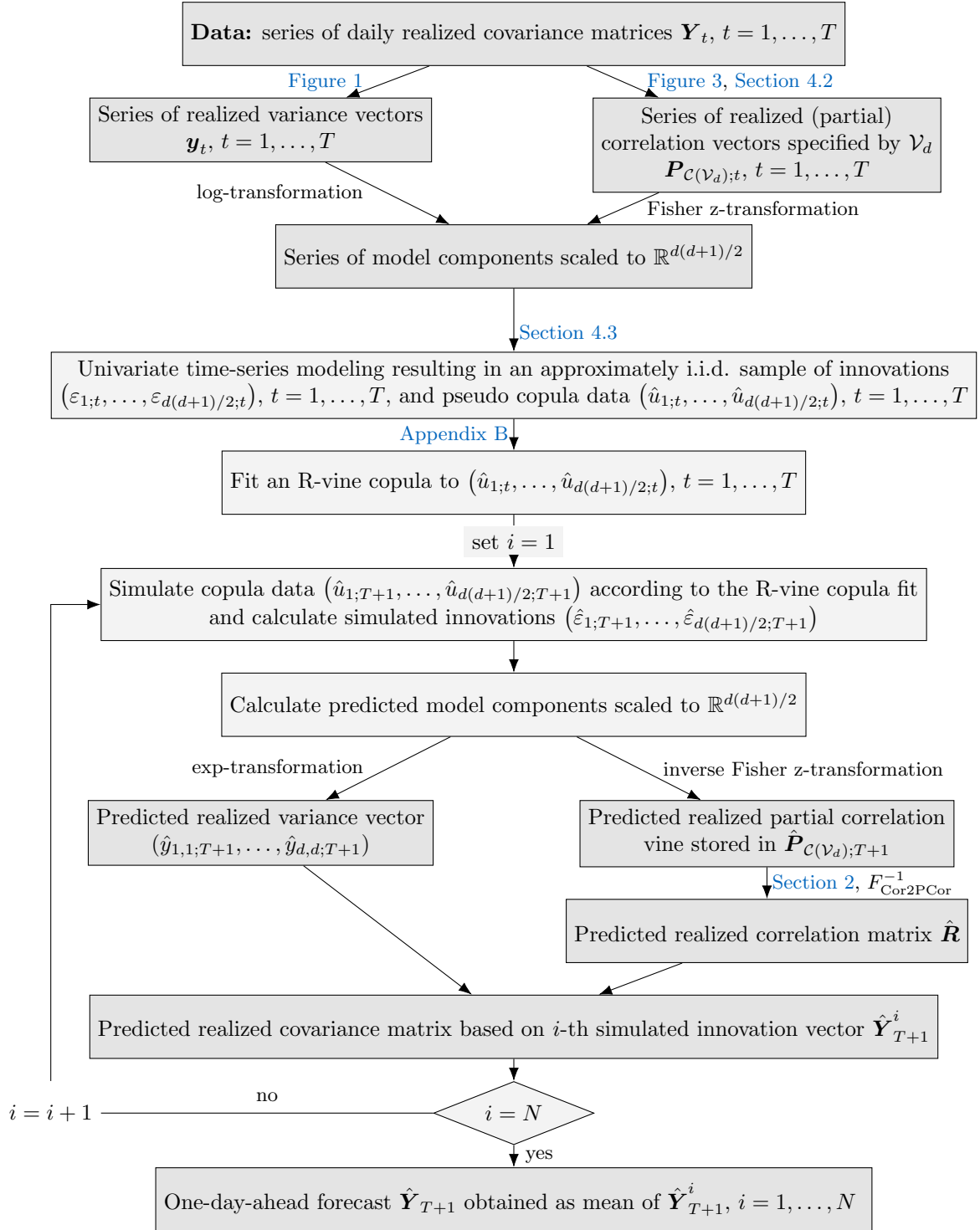
Figure 6: Modeling and forecasting approach using partial correlation vine based data transformation of the series of realized covariance matrices in step (S1) and an R-vine copula based time-series model in step (S2). The one-day-ahead forecast $\hat{\boldsymbol{Y}}_{T+1}$ is obtained as mean of N simulation based matrix forecasts.
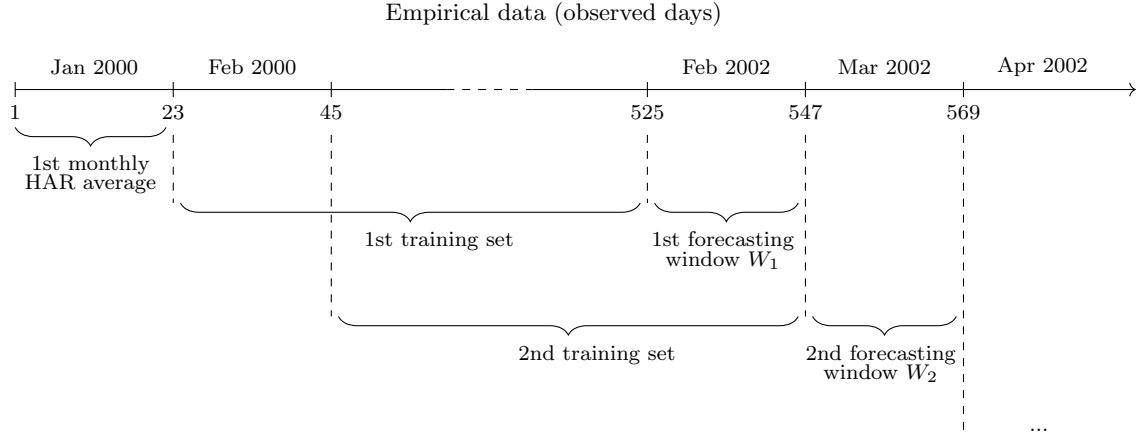
Empirical data (observed days)

Jan 2000   Feb 2000   Feb 2002   Mar 2002   Apr 2002

1   23   45   525   547   569

1st monthly
HAR average

1st training set   1st forecasting
                        window W_1

2nd training set   2nd forecasting
                        window W_2

...

Figure 7: Moving window approach illustrated for the considered real data example.

In Figure 8, the first trees of the R-vine structures selected in each of the 75 time windows are shown indicating the included model components by a black square. In the first row, empirical means of the realized standard correlations are considered. In the second and third row, exponentially weighted moving averages based on $\lambda = 0.995$ and $\lambda = 0.98$, respectively, are used. While in case of empirical means all days of the two training years are of equal weight, for $\lambda = 0.995$ and $\lambda = 0.98$ the six most recent months and one and a half months, respectively, already contribute half of the information for average calculation. Thus, in the latter case R-vine structure selection is most sensitive to market changes resulting in more frequent variations of the selected model components. Changes e.g. for the prediction months in mid 2004 or at the beginning of the financial crisis are observed earliest. Nevertheless, for all three setups the selected first tree is quite stable and we may identify three distinct periods: February 2002–August 2006, September 2006–July 2007 and August 2007–July 2008. For these periods, Figure 9 illustrates the first tree $\mathcal{T}_1$ of the predominantly chosen R-vine structures. Until August 2006, pairwise correlations between the log-returns including Citigroup (C) and General Electric (GE) seem to be most pronounced. While C plays a key role within the financial sector, GE as a diversified industrial corporation connects the representatives of the financial sector with the two non-financial stocks. During the period from September 2006 to July 2007 C becomes the first root node in a C-vine, i.e. the node in $\mathcal{T}_1$ with the highest possible number of edges attached to it. At the beginning of the financial crisis in August 2007, the correlations between JP Morgan (JPM) and the other market participants seem to tighten. This results in a predominantly chosen R-vine structure, where except for GE all pairwise correlations including JPM are modeled. Note that in 2007 JPM replaced C as the biggest US-bank in terms of revenues.

To conclude, selecting the R-vine structure for data transformation as proposed in Section 4.2 gives interesting insights into market activities over time. In addition, we already know about the resulting inhomogeneous data complexity of the corresponding time-series, which will be further analyzed in the next section. There, the R-vine structures will be selected using EWMA based edge weights with $\lambda = 0.995$. Recall, however, that any R-vine structure could be used for data transformation. To demonstrate the general adequacy of the partial correlation vine data transformation approach irrespective of the R-vine structure used for data transformation, we will consider two alternative ways of R-vine structure selection as well. First, we reverse the idea of inducing model parsimony and select for each time window a C-vine, where in each tree level the root node induces the on average lowest correlation strength. Thus, the effect of decreasing data complexity as for the proposed R-vine structure selection should be eliminated. Second, an R-vine structure on 6 elements is randomly sampled in each time window (Joe et al. 2011).
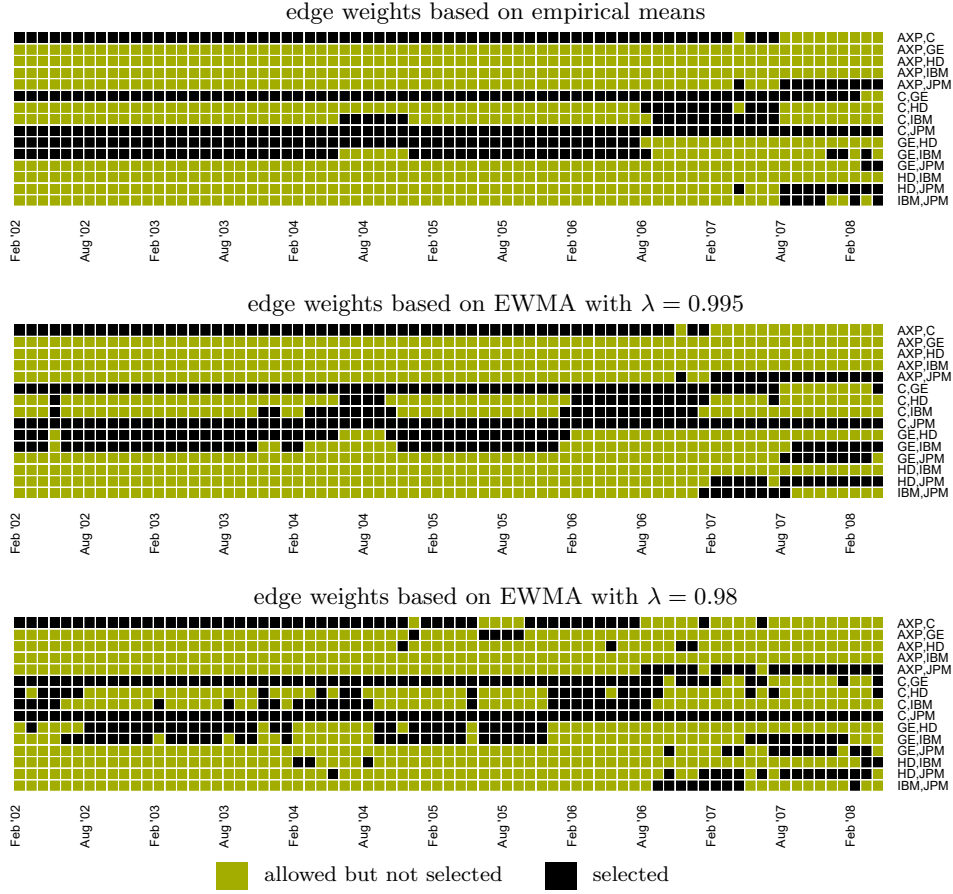
Figure 8: Illustration of the dynamically changing R-vine structures used for transforming the series of realized correlation matrices in each of the 75 time windows. The horizontal time axis states the prediction month. Model components selected in tree $\mathcal{T}_1$ are indicated by black squares. Green squares indicate selectable components allowed by the proximity condition (which does not trigger in $\mathcal{T}_1$). In the first panel, the average correlations used for R-vine structure selection are the empirical means of the training set data. In the second and third row, they are exponentially weighted moving averages with $\lambda = 0.995$ and $\lambda = 0.98$, respectively.
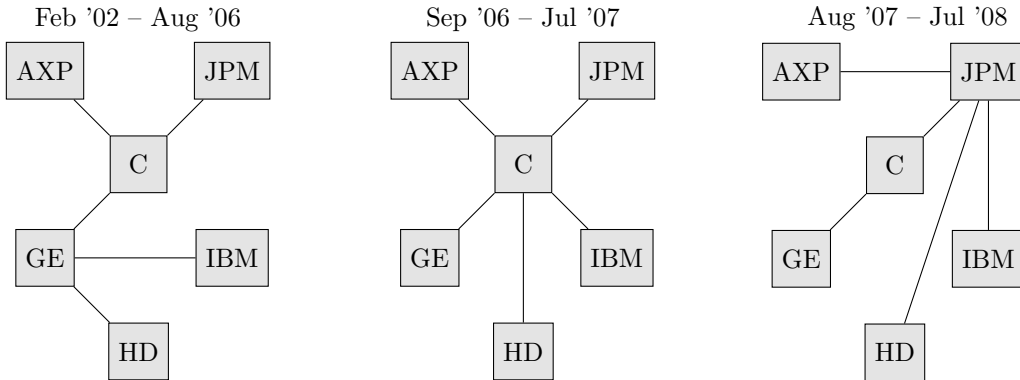

Figure 9: Predominantly selected first tree of the R-vine structure used for data transformation during the periods February 2002 to August 2006, September 2006 to July 2007 and August 2007 to July 2008.

21

For the Cholesky decomposition, the model components depend on the ordering of the assets. However, as opposed to the data transformation based on partial correlation vines there is no justifiable rule to decide 'on the fly' for a specific order. Thus, the ordering has to be set upfront. In this sense, the Cholesky decomposition based data transformation is static. As the only way to proceed, Brechmann et al. (2018) test all possible permutations for the underlying data. Coincidently, the alphabetic ordering of the six stocks performs best. We therefore, choose the latter for all time windows.

5.3 Step (S2): Multivariate time-series modeling

As explained in Section 4.3, to the sample of model components obtained after transforming the series of realized covariance matrices, marginal time-series models need to be applied first.

Marginal time-series modeling

Given the proposed R-vine structure selection method, we know that with increasing tree level the data complexity decreases such that less elaborate time-series models might already be sufficient for accurate in-sample estimation and out-of-sample forecasting. To support this presumption, for each period within the moving window approach time-series models of different complexity are fitted to the log-transformed realized variance time-series and to the Fisher z-transformed realized (partial) correlations specified by the R-vine structure found in Section 4.2. For comparison, also the time-series specified by the C-vine, of which in each tree level the root node induces the on average lowest correlation strength, are investigated.

Besides simply considering the mean value over time, basic univariate HAR and ARFIMA models as well as HAR and ARFIMA models including a GARCH(1,1) component with normal innovations (abbreviated as HN and AN) and with SGED innovations (abbreviated as HSGED and ASGED) are fitted. To evaluate the statistical precision we use the root mean squared error (RMSE), which according to Patton (2011) is robust to noise in the volatility proxies. Table 2 shows the out-of-sample RMSE for all time-series model components under consideration. In each row, the set of superior models based on the MCS approach of Hansen et al. (2011) with a confidence level of 10% is highlighted in gray. The model with the lowest RMSE, which is the last one that would be rejected from the model confidence set, is highlighted in bold. In general, ARFIMA based models show a superior prediction performance with respect to the RMSE criterion. However, especially within the variations of the two base models the RMSE values often are very close to each other. For the realized variance time-series and the realized standard correlation time-series in $\mathcal{T}_1$ of the selected R-vine structure, the best model usually includes a GARCH(1,1) augmentation. For tree level $\mathcal{T}_2$ and $\mathcal{T}_3$, there is a shift to basic ARFIMA models, while for tree level $\mathcal{T}_4$ and $\mathcal{T}_5$ even simply using the mean realized partial correlation value as forecast is included in the model confidence set at a confidence level of 10%. This confirms the presumption that given the proposed R-vine structure selection method with increasing tree level more parsimonious time-series models already are sufficient. This hierarchical pattern is not observed for the considered C-vine. Here, base models including a GARCH(1,1) augmentation with normal or SGED innovations most often would be the last ones to be eliminated from the model confidence set. In particular, a simple mean forecast clearly is insufficient even in high tree levels. Similar results are detected for the Cholesky elements and are given in Table 5 in Appendix C.

In the following, we consider two groups of models. One including only HAR based time-series models and the other including only ARFIMA based models. Given the above findings within the partial correlation vine data transformation approach, we use HN and AN models, respectively, for all components in case that a C-vine or a randomly sampled R-vine structure is taken for data transformation. Likewise, we proceed for the Cholesky decomposition based

Table 2: RMSE with respect to the complete out-of-sample forecasting horizon (1632 days) for the model components in the partial correlation vine data transformation approach. Two different R-vine structures for data transformation are considered. The set of superior models according to the MCS approach at a confidence level of 10% is highlighted in gray. The lowest RMSE is highlighted in bold.

		mean	HAR	HN	HSGED	ARFIMA	AN	ASGED
	AXP	1.0429	0.4711	0.4715	0.4719	0.4684	**0.4668**	0.4680
	C	1.0154	0.4469	0.4479	0.4510	0.4465	**0.4455**	0.4483
	GE	0.8105	0.4634	0.4632	0.4647	0.4627	**0.4625**	0.4627
	HD	0.7766	0.4554	0.4557	0.4568	**0.4540**	0.4543	0.4543
	IBM	0.7242	0.4320	0.4322	0.4323	**0.4317**	0.4331	0.4327
	JPM	1.0137	0.4653	0.4647	0.4671	0.4652	**0.4641**	0.4655
R-vine selection (Section 4.2)	AXP,C	0.2183	0.1572	0.1573	0.1576	0.1568	**0.1568**	0.1571
	C,GE	0.1984	0.1531	0.1531	0.1530	0.1526	0.1526	**0.1524**
	C,HD	0.1857	0.1519	0.1520	0.1520	0.1515	0.1516	**0.1512**
	C,JPM	0.2149	0.1619	0.1620	0.1620	0.1615	**0.1615**	0.1615
	GE,IBM	0.1914	**0.1489**	0.1490	0.1490	0.1490	0.1490	0.1490
	AXP,GE;C	0.1395	0.1317	0.1317	0.1316	**0.1313**	0.1313	0.1313
	AXP,JPM;C	0.1364	**0.1295**	0.1295	0.1296	0.1297	0.1297	0.1297
	C,IBM;GE	0.1340	0.1271	0.1271	0.1272	**0.1270**	0.1270	0.1270
	GE,HD;C	0.1384	0.1300	0.1301	0.1301	**0.1292**	0.1292	0.1292
	AXP,IBM;C,GE	0.1246	0.1237	0.1237	0.1237	**0.1231**	0.1231	0.1232
	GE,JPM;AXP,C	0.1211	0.1196	0.1196	0.1196	**0.1191**	0.1191	**0.1191**
	HD,IBM;C,GE	0.1260	0.1253	0.1254	0.1254	**0.1250**	0.1250	0.1251
	AXP,HD;C,GE,IBM	0.1175	0.1171	0.1171	0.1172	0.1168	**0.1168**	0.1170
	IBM,JPM;AXP,C,GE	0.1237	0.1227	0.1227	0.1227	0.1225	**0.1223**	0.1225
	HD,JPM;AXP,C,GE,IBM	0.1177	**0.1175**	0.1175	0.1175	0.1178	0.1177	0.1178
C-vine	AXP,HD	0.1873	0.1547	0.1547	0.1549	**0.1538**	0.1538	0.1539
	C,HD	0.1857	0.1519	0.1520	0.1520	0.1515	0.1516	**0.1512**
	GE,HD	0.1856	0.1517	0.1517	0.1517	0.1509	**0.1509**	0.1509
	HD,IBM	0.1731	0.1488	0.1489	0.1489	0.1484	0.1484	**0.1483**
	HD,JPM	0.1804	0.1532	0.1533	0.1533	0.1525	**0.1523**	0.1525
	AXP,IBM;HD	0.1491	0.1349	0.1349	0.1349	0.1345	**0.1344**	0.1345
	C,IBM;HD	0.1502	0.1320	0.1320	0.1320	0.1317	0.1317	**0.1315**
	GE,IBM;HD	0.1574	0.1352	0.1352	0.1353	**0.1353**	0.1354	0.1355
	IBM,JPM;HD	0.1486	0.1382	0.1382	0.1383	0.1376	0.1376	**0.1376**
	AXP,GE;HD,IBM	0.1387	0.1303	0.1303	0.1302	0.1299	0.1300	**0.1299**
	C,GE;HD,IBM	0.1371	0.1270	0.1270	0.1268	0.1270	0.1270	**0.1268**
	GE,JPM;HD,IBM	0.1312	0.1251	**0.1251**	0.1251	0.1251	0.1251	0.1252
	AXP,C;GE,HD,IBM	0.1496	0.1283	0.1283	0.1285	0.1281	**0.1281**	0.1282
	AXP,JPM;GE,HD,IBM	0.1526	0.1321	0.1321	0.1321	0.1317	**0.1316**	0.1316
	C,JPM;AXP,GE,HD,IBM	0.1398	0.1247	0.1247	0.1248	0.1241	**0.1241**	0.1242

model. In case of R-vine structure selection according to Section 4.2, we stepwise increase model parsimony. For model components corresponding to the realized variance and realized standard correlation time-series in $\mathcal{T}_1$, we use HN and AN models, respectively. For the ones in tree level $\mathcal{T}_2$ and $\mathcal{T}_3$ we apply basic HAR and ARFIMA models, respectively. For components in $\mathcal{T}_4$ and $\mathcal{T}_5$ we consider in one setting basic HAR and ARFIMA models, respectively, and take in another setting simply the mean value of the underlying training set as forecast.

Dependence modeling

Now, for each time window interest is in the cross-sectional dependence between the model components. Since dependencies between stocks are expected to be most pronounced during financial turmoil, we consider as an example the time window from July 2006 to July 2008. Based on the specified time-series models, the sample of innovations is obtained and transformed to pseudo copula data (Section 4.3). Figure 10 shows the resulting data based on R-vine structure selection according to Section 4.2 and HAR based time-series modeling. It illustrates the corresponding histograms on its diagonal, pairwise contour plots with standard normal margins in the lower left corner and pairs plots with corresponding Kendall's τ values in the upper right corner. Only dependencies between model components corresponding to realized variances (last six components) and realized standard correlations (first 5 components) are significant with Kendall's τ values ranging from 0.2 to 0.5. Dependencies including components, which correspond to partial correlations, are rather small and close to zero for higher tree levels. Based on these findings, we subsequently consider five different R-vine copula settings. First, independence for all pairs is assumed. Second, a 21-dimensional R-vine copula is fitted to capture dependence between all model components. Third, a reduced

Figure 10: Exploratory data analysis for the pairwise dependencies of the 21-dimensional pseudo copula data estimated for the period July 2006 to July 2008 using the proposed method for R-vine structure selection (Section 4.2) and HAR based time-series modeling. Pairwise contour plots with normalized margins, histograms and pairs plots with empirical Kendall's τ values are shown. The first five components PCV1 to PCV5 correspond to realized standard correlations in $\mathcal{T}_1$, components PCV6 to PCV9 correspond to realized first order partial correlations in $\mathcal{T}_2$, etc. Variables PCV16 to PCV21 correspond to realized variances.

24

structured dependence is imposed, where a 11-dimensional R-vine copula is fitted only to the components corresponding to realized variances and realized standard correlations. The components corresponding to realized partial correlations are assumed to be independent. Both in case of full and reduced structured R-vine copula based dependence modeling, we allow as a first setting the pair-copulas to stem from various copula families such as Clayton, Gumbel, Frank, etc. including their reflected forms. Thus, possible asymmetric and nonlinear dependence patterns can be detected. Given the primarily elliptical shapes in Figure 10 we also consider an R-vine copula exclusively built from bivariate (conditional) Gaussian copulas, i.e. a Gaussian vine. Except for the structured dependence, the same settings for the copula models are taken in the Cholesky decomposition based benchmarks. To fit an R-vine copula model we rely on the R-package `VineCopula` (Schepsmeier et al. 2017).

5.4   Forecasting performance

Based on the above model specifications, one-day-ahead forecasts as described in Section 4.4 are obtained. In addition to the 36 data transformation based prediction models, we consider three naive benchmarks. First, $\hat{\boldsymbol{Y}}_{T+1}$ is set to the realized covariance matrix at time point T, i.e. $\hat{\boldsymbol{Y}}_{T+1} = \boldsymbol{Y}_T$. Second, $\hat{\boldsymbol{Y}}_{T+1}$ is calculated as the equally weighted average of the realized covariance matrices in the corresponding training set. Third, $\hat{\boldsymbol{Y}}_{T+1}$ is obtained as an exponentially weighted moving average, i.e. in our setup $\hat{\boldsymbol{Y}}_{T+1} = \lambda \hat{\boldsymbol{Y}}_T + (1 - \lambda) \boldsymbol{Y}_T$, where the smoothing parameter λ is set to 0.94 as commonly suggested in the framework of a RiskMetrics approach (Morgan 1996).

Out-of-sample forecasting precision

To illustrate that the proposed forecasting approach is on target, Figure 11 shows for the realized variance time-series of JPM (top panel), the realized covariance time-series of C and JPM (mid panel) as well as IBM and JPM (bottom panel) the historical time-series from January 2002 until July 2008 together with the one-day-ahead forecasts based on the R-vine structure selected according to Section 4.2, ARFIMA based time-series modeling and a 21-dimensional Gaussian vine for dependence modeling. Results for all other realized variances and covariance pairs are similar and given in Figure 14 in Appendix C. The trends in all time-series including high short-term peaks are well detected and modeled. Distances between historical extreme peaks and corresponding forecasts are large. This finding holds true for all prediction models and is due to the high volatility of the realized variances and covariances. The predicted time-series incorporate smoothed long-term information of historical data and thus, are more stable.

To evaluate the statistical precision of the matrix forecasts, Table 3 summarizes for all considered models the RMSE based on the Frobenius norm between the realized and the predicted covariance matrices. This loss function satisfies the conditions in Laurent et al. (2013) for consistent model ranking. In the right column, the RMSE based on bias corrected (bc) matrix forecasts are shown. We use historical data over the period of one year for level correction reducing the out-of-sample forecasting horizon to 1368 days. As in the previous analysis of the single model components, ARFIMA based models in general have smaller RMSE values compared to HAR based models. All models using partial correlation vine based data transformation and full dependence modeling exhibit a smaller RMSE than Cholesky decomposition based models and show very similar performance among each other. This confirms that any R-vine structure can be used for data transformation in step (S1) of the model approach. Among the partial correlation vine data transformation based models those with a C-vine structure used for data transformation have the highest RMSE. Recall that by construction more complex data features are induced even for high tree levels. For C-vine and random R-vine structures, time-series modeling in step (S2) with independent components and reduced structured dependence between components is clearly improved by
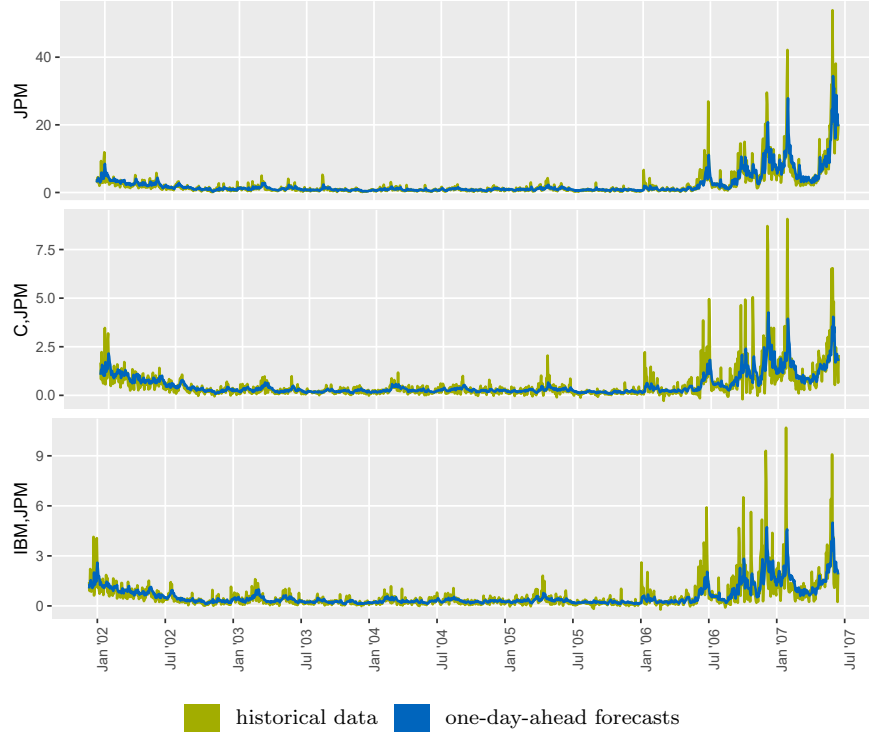
Figure 11: Daily realized variance time-series for JPM (1st row) and daily realized covariance time-series (2nd and 3rd row) together with the time-series of the corresponding daily forecasts based on the partial correlation vine data transformation approach with R-vine structure selected according to Section 4.2, ARFIMA based time-series modeling and a 21-dimensional Gaussian vine for dependence modeling.

models, which capture dependence between all model components. Here, the decreasing data complexity does not trigger. However, for the R-vine structure selected according to Section 4.2 the performance in case of reduced structured dependence is only slightly improved by full dependence modeling. Thus, also the dependence between the model components allows for model parsimony. In general, using a Gaussian vine for dependence modeling between the model components shows comparable results as using more elaborate copulas allowing e.g. for tail-dependence. All discussed prediction models clearly show superior results as compared to the naive benchmarks. Bias correction in step (S3) slightly improves results while maintaining the above observations among the different models.

To test the statistical significance of the results, we apply the MCS approach of Hansen et al. (2011). Based on the above findings, we restrict the analysis to models using a Gaussian vine for dependence modeling between the model components. Only in case of R-vine structure selection according to Section 4.2 we consider reduced structured dependence modeling in addition to the full one. Figure 12 shows for each half-year period of the out-of-sample horizon the set of superior models (indicated by a gray dot), which contains the best model at a confidence level of 10%. A blue triangle and an orange cross indicate the last and the next model, respectively, that would be eliminated. For almost all periods, all models are selected at the given confidence level showing very close performance of all models. Most often, HAR based models would be eliminated next, while ARFIMA based models usually would be the last ones to be eliminated from the set of superior models. In three out of eleven periods, the ARFIMA-Cholesky model has the smallest RMSE based on the Frobenius norm and therefore automatically would be the last model to be eliminated. All ARFIMA and partial correlation vine data transformation based models show rather robust performance over the out-of-sample forecasting horizon. Especially, the models based on R-vine structure selection according to Section 4.2 usually are the ones to be eliminated last from the MCS, i.e. having the smallest loss.

26

Table 3: RMSE based on the Frobenius norm between the realized and predicted correlation matrices with respect to the complete out-of-sample forecasting horizon (1368 days) for all models.

Marginals	Data transformation based on		R-vine copula assumed for transformed data	RMSE	RMSE bc
ARFIMA based	PCV	R-vine selection ([Section 4.2](#))	independence	6.6314	6.6045
			full all	6.5725	6.5632
			full Gauss	6.5685	6.5619
			structured all	6.5858	6.5740
			structured Gauss	6.5864	6.5742
		C-vine	independence	6.7076	6.6733
			full all	6.5968	6.5854
			full Gauss	6.5967	6.5860
			structured all	6.6436	6.6189
			structured Gauss	6.6410	6.6166
		random R-vine	independence	6.6694	6.6393
			full all	6.5826	6.5746
			full Gauss	6.5886	6.5810
			structured all	6.6155	6.5968
			structured Gauss	6.6138	6.5954
		Cholesky	independence	6.6732	6.6437
			all	6.6121	6.6001
			Gauss	6.6193	6.6075
HAR based	PCV	R-vine selection ([Section 4.2](#))	independence	6.7218	6.6566
			full all	6.6332	6.5998
			full Gauss	6.6313	6.5962
			structured all	6.6544	6.6146
			structured Gauss	6.6522	6.6122
		C-vine	independence	6.7900	6.7085
			full all	6.6574	6.6153
			full Gauss	6.6575	6.6158
			structured all	6.7117	6.6480
			structured Gauss	6.7094	6.6453
		random R-vine	independence	6.7527	6.6830
			full all	6.6432	6.6117
			full Gauss	6.6474	6.6158
			structured all	6.6821	6.6334
			structured Gauss	6.6796	6.6310
		Cholesky	independence	6.7400	6.6866
			all	6.6863	6.6621
			Gauss	6.6841	6.6603
mean over training set				12.0894	
previous day				7.2937	
EWMA with $\lambda = 0.94$				7.8790	

Figure 12: Model confidence sets of Hansen et al. (2011) with confidence level 10% for all half-year periods of the out-of-sample forecasting horizon. Gray dots indicate selected models, blue triangles and orange crosses indicate the last and the next model, respectively, that would be eliminated from the set of superior models.

Mean-variance trade-off in portfolio optimization

For additional economic evaluation of the forecasts, we construct portfolios based on each prediction model, which are mean-variance efficient. For a risk-averse investor we assume a quadratic utility function. Then, the problem to maximize the utility is reduced to finding the asset weights $\boldsymbol{w}$, which minimize the portfolio volatility σ_{p} based on a fixed target expected return μ_{p} (Markowitz 1952). The optimal portfolio is obtained by solving the quadratic problem

$$\min_{\boldsymbol{w}_{t+1}} \boldsymbol{w}_{t+1}'\hat{\boldsymbol{\Sigma}}_{t+1}\boldsymbol{w}_{t+1} \quad \text{s.t.} \quad \boldsymbol{w}_{t+1}'\mathbb{E}\left[\boldsymbol{r}_{t+1}|\mathcal{F}_t\right] = \mu_{\mathrm{p}} \quad \text{and} \quad \boldsymbol{w}_{t+1}'\mathbf{1}_d = 1,$$

where $\boldsymbol{w}_{t+1}$ is the $d \times 1$ vector of portfolio weights chosen at day t for $t+1$, $\mathbf{1}_d$ is a $d \times 1$ vector of ones, μ_{p} is the daily target expected return and $\hat{\boldsymbol{\Sigma}}_{t+1}$ is the conditional (with respect to the information set) covariance forecast at day t for $t + 1$. The latter corresponds to the realized covariance forecasts $\hat{\boldsymbol{Y}}_{t+1}$.

For each prediction model, we solve the above optimization problem for a daily target return μ_{p} for all 1368 days in the out-of-sample horizon. Based on the optimal portfolio weights $\boldsymbol{w}_t$ for day t ($t = 1, \ldots, 1368$) the expected risk in terms of standard deviation, $\sqrt{\boldsymbol{w}_t'\hat{\boldsymbol{Y}}_t\boldsymbol{w}_t}$, corresponding to the target expected return μ_{p} can be calculated. Taking the averages over the forecasting horizon and repeating the procedure for a grid of target returns, results in an average efficient frontier for each prediction model. To obtain an average oracle efficient frontier, the true realized covariance matrices for each day t are used. Figure 13 shows the efficient frontiers for the considered HAR and ARFIMA based prediction models. All partial correlation vine data transformation based models show a clear improvement in terms of the expected mean-variance trade-off compared to the two Cholesky decomposition based prediction models.

To validate this observation in an out-of-sample setting we calculate for each prediction model based on the corresponding optimal portfolio weights $\boldsymbol{w}_{t+1}$ estimated at day t for $t+1$ ($t = 0, \ldots, 1367$) the ex-post realized portfolio return $r_{\mathrm{p},t+1} = \boldsymbol{w}_{t+1}'\boldsymbol{r}_{t+1}$ and the ex-post realized portfolio volatility $\sigma_{\mathrm{p},t+1} = \sqrt{\boldsymbol{w}_{t+1}'\boldsymbol{Y}_{t+1}\boldsymbol{w}_{t+1}}$. Here, $\boldsymbol{r}_{t+1}$ and $\boldsymbol{Y}_{t+1}$ are the true returns and the true covariance matrix, respectively, realized at day $t + 1$. Given a small enough grid of target returns, we are able to obtain for each prediction model the series
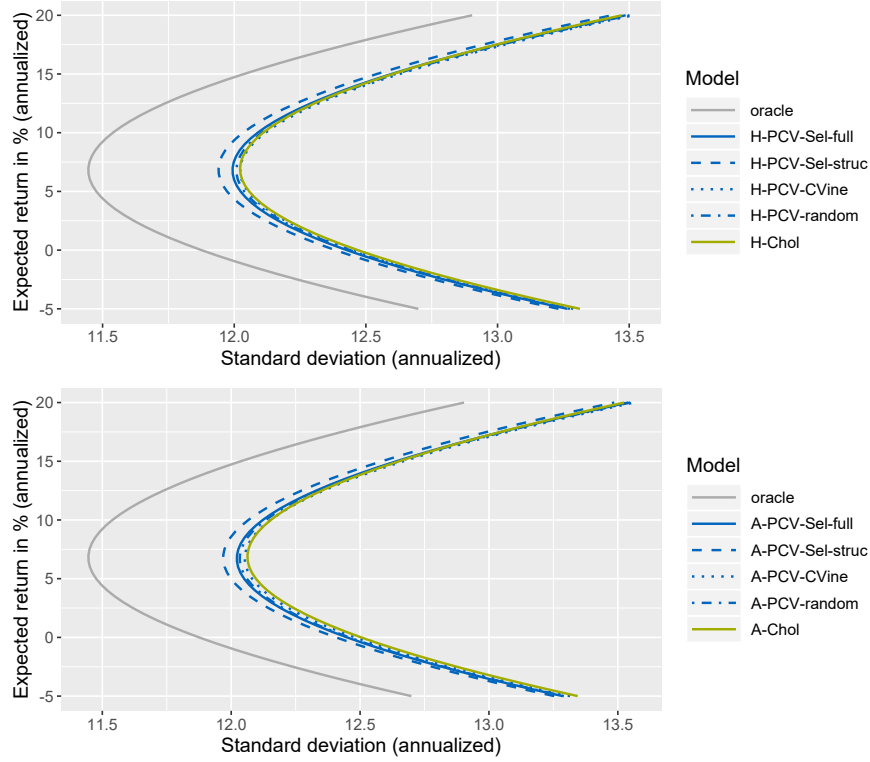
28

Figure 13: Efficient frontier for each HAR and ARFIMA based prediction model plotting the expected return versus its corresponding risk in terms of standard deviation. The curves are obtained as averages over the out-of-sample horizon (1368 days).

of ex-post portfolio standard deviation $\sigma_{p,t+1}$, $t = 0, \ldots, 1367$, corresponding to a certain average ex-post realized return. For an average annualized ex-post realized portfolio return of approximately 7.5%, 10%, 12% and 15%, Table 4 shows the average annualized ex-post realized portfolio standard deviation for each prediction model. The set of models, which includes the model with the lowest standard deviation at a confidence level of 10% based on the MCS approach of Hansen et al. (2011), is highlighted in gray. The model with the lowest loss (deviation from zero) is highlighted in bold. In general, HAR based prediction models perform better than their ARFIMA based counterparts. In the ex-post analysis, all ARFIMA based partial correlation vine data transformation based models have the highest average standard deviation. This confirms the often seen phenomenon that models with the lowest statistical loss do not necessarily show superior results in economical applications (Laurent et al. 2013). The HAR based model with R-vine structure selected according to Section 4.2 and with reduced structured dependence among the model components is the best model at a confidence level of 10% for all considered annualized ex-post realized portfolio returns. Comparing the average ex-post realized standard deviations of the HAR based prediction models, further demonstrates the strength of the proposed methodology irrespective of the R-vine structure used for data transformation.

Table 4: Annualized average ex-post standard deviation corresponding to four levels of annualized ex-post realized return. The set of models, which includes the one with the smallest standard deviation at a confidence level of 10%, are highlighted in gray. The last model to be eliminated is highlighted bold.

Model	Realized return in % (annualized)			
	7.5	10	12.5	15
A-PCV-Sel-full	12.5217	12.9545	13.5492	14.2832
A-PCV-Sel-struc	12.5055	12.9254	13.5086	14.2317
A-PCV-CVine	12.4981	12.9269	13.5252	14.2722
A-PCV-random	12.4949	12.9151	13.5027	14.2366
A-Chol	12.4616	12.8641	13.4352	14.1510
H-PCV-Sel-full	12.4754	12.8748	13.4359	14.1363
H-PCV-Sel-struc	**12.4588**	**12.8447**	**13.3937**	**14.0796**
H-PCV-CVine	12.4595	12.8606	13.4293	14.1429
H-PCV-random	12.4680	12.8680	13.4349	14.1413
H-Chol	12.4718	12.8729	13.4396	14.1475

6 Discussion

In this paper, we introduce a novel approach to model and forecast time-series of realized covariance matrices. Realized variances and realized correlation matrices are jointly modeled. We address the challenge of generating symmetric and positive semi-definite correlation matrix forecasts by introducing partial correlation vines as a tool to transform the series of realized correlation matrices. Along with a real data example we explore in detail the benefits of the proposed methodology as compared to Cholesky decomposition based competitor models. Given the large number of R-vine structures for data transformation, we propose an R-vine structure selection method, which exclusively relies on historical information of the underlying data. This procedure allows the R-vine structure to dynamically change over time and therewith to adapt to market changes. The selection method is motivated by the practical interpretation of the model components, which are proxies for the conditional variances and conditional correlations corresponding to a daily log-return series. High average correlation strengths are captured in lower tree levels of the R-vine structure leaving higher order realized partial correlation time-series, for which parsimonious univariate time-series modeling is sufficient. For the latter also dependence is negligible allowing for dimension reduction in the multivariate time-series model. The forecasting performance both in terms of statistical precision and in an economic evaluation, where ex-post realizations of mean-variance efficient portfolios are investigated, shows very good and in several settings even statistically significant superior prediction capability compared to the Cholesky decomposition based benchmark models. Given the excellent prediction power of the latter often demonstrated in literature, these findings provide strong evidence for the use of the partial correlation vine data transformation approach in practice.

Acknowledgments

References

Andersen, T. G. and Bollerslev, T. (1997). Heterogeneous information arrivals and return volatility dynamics: Uncovering the long-run in high frequency returns. *The Fournal of Finance*, 52(3):975–1005.

Andersen, T. G., Bollerslev, T., Christoffersen, P. F., and Diebold, F. X. (2006). Volatility and correlation forecasting. *Handbook of Economic Forecasting*, 1:777–878.

Andersen, T. G., Bollerslev, T., Diebold, F. X., and Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*, 61(1):43–76.

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

Anderson, T. W. (1958). *An introduction to multivariate statistical analysis*. Wiley New York.

Bai, X., Russell, J. R., and Tiao, G. C. (2003). Kurtosis of garch and stochastic volatility models with non-normal innovations. *Journal of Econometrics*, 114(2):349–360.

Baillie, R. T., Bollerslev, T., and Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 74(1):3–30.

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.

Bedford, T. and Cooke, R. M. (2002). Vines: A new graphical model for dependent random variables. *Annals of Statistics*, pages 1031–1068.

Bernstein, D. S. (2005). *Matrix mathematics: Theory, facts, and formulas with application to linear systems theory*, volume 41. Princeton university press Princeton.

Brechmann, E. C., Heiden, M., and Okhrin, Y. (2018). A multivariate volatility vine copula model. *Econometric Reviews*, 37(4):281–308.

Brechmann, E. C. and Joe, H. (2014). Parsimonious parameterization of correlation matrices using truncated vines and factor analysis. *Computational Statistics & Data Analysis*, 77:233–251.

Brechmann, E. C. and Joe, H. (2015). Truncation of vine copulas using fit indices. *Journal of Multivariate Analysis*, 138:19–33.

Cappiello, L., Engle, R. F., and Sheppard, K. (2006). Asymmetric dynamics in the correlations of global equity and bond returns. *Journal of Financial Econometrics*, 4(4):537–572.

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

Christensen, K., Kinnebrock, S., and Podolskij, M. (2010). Pre-averaging estimators of the ex-post covariance matrix in noisy diffusion models with non-synchronous data. *Journal of Econometrics*, 159(1):116–133.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.

Corsi, F., Mittnik, S., Pigorsch, C., and Pigorsch, U. (2008). The volatility of realized volatility. *Econometric Reviews*, 27(1-3):46–78.

Czado, C. (2010). Pair-copula constructions of multivariate copulas. In *Copula Theory and its Applications*, pages 93–109. Springer.

Doléans-Dade, C. and Meyer, P.-A. (1970). Intégrales stochastiques par rapport aux martingales locales. In *Séminaire de Probabilités IV Université de Strasbourg*, pages 77–107. Springer.

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

Fernández, C. and Steel, M. F. (1998). On bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, 93(441):359–371.

Gençay, R., Dacorogna, M., Muller, U. A., Pictet, O., and Olsen, R. (2001). *An introduction to high-frequency finance*. Academic press.

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.

Gouriéroux, C., Jasiak, J., and Sufana, R. (2009). The wishart autoregressive process of multivariate stochastic volatility. *Journal of Econometrics*, 150(2):167–181.

Halbleib, R. and Voev, V. (2014). Forecasting covariance matrices: A mixed approach. *Journal of Financial Econometrics*, 14(2):383–417.

Hansen, P. R., Huang, Z., and Shek, H. H. (2012). Realized GARCH: A joint model for returns and realized measures of volatility. *Journal of Applied Econometrics*, 27(6):877–906.

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

Hayashi, T., Yoshida, N., et al. (2005). On covariance estimation of non-synchronously observed diffusion processes. *Bernoulli*, 11(2):359–379.

Hobæk Haff, I., Aas, K., and Frigessi, A. (2010). On the simplified pair-copula construction-simply useful or too simplistic? *Journal of Multivariate Analysis*, 101(5):1296–1310.

Jacod, J. (1994). Limit of random measures associated with the increments of a brownian semimartingale. *preprint*, 120:155–162.

Joe, H. (1996). Families of m-variate distributions with given margins and $m(m-1)/2$ bivariate dependence parameters. In Rüschendorf, L., Schweizer, B., and Taylor, M. D., editors, *Distributions with fixed marginals and related topics*, pages 120–141. Institute of Mathematical Statistics, Hayward.

Joe, H. (1997). *Multivariate models and dependence concepts*. Chapman and Hall, London.

Joe, H. (2005). Asymptotic efficiency of the two-stage estimation method for copula-based models. *Journal of Multivariate Analysis*, 94(2):401–419.

Joe, H. (2014). Dependence Modeling with Copulas, Chapman Hall/CRC. *Published June/July*.

Joe, H., Cooke, R. M., and Kurowicka, D. (2011). Regular vines: generation algorithm and number of equivalence classes. In *Dependence Modeling: Vine Copula Handbook*, pages 219–231. World Scientific.

Joe, H. and Xu, J. (1996). The estimation method of inference functions for margins for multivariate models. *Journal of Multivariate Analysis*, 166.

Katoh, N., Ibaraki, T., and Mine, H. (1981). An algorithm for finding k minimum spanning trees. *SIAM Journal on Computing*, 10(2):247–255.

Kurowicka, D. and Cooke, R. (2003). A parameterization of positive definite matrices in terms of partial correlation vines. *Linear Algebra and its Applications*, 372:225–251.

Kurowicka, D. and Cooke, R. (2006). Completion problem with partial correlation vines. *Linear Algebra and its Applications*, 418(1):188–200.

Kurowicka, D. and Joe, H. (2011). *Dependence modeling: Vine Copula Handbook*. World Scientific.

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

Lewandowski, D., Kurowicka, D., and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis*, 100(9):1989–2001.

Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1):77–91.

Morales Napoles, O., Cooke, R. M., and Kurowicka, D. (2010). About the number of vines and regular vines on n nodes.

Morgan, J. P. (1996). Riskmetrics technical document.

Patton, A. J. (2011). Volatility forecast comparison using imperfect volatility proxies. *Journal of Econometrics*, 160(1):246–256.

Poignard, B. (2017). *New approaches for high-dimensional multivariate GARCH models*. PhD thesis, PSL Research University.

Schepsmeier, U., Stoeber, J., Brechmann, E., and Gräler, B. (2017). VineCopula: Statistical inference of vine copulas. *R Package Version 1.7/r66*.

Sklar, A. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Publ. Inst. Stat. Université Paris 8.

Stoeber, J., Joe, H., and Czado, C. (2013). Simplified pair copula constructionslimitations and extensions. *Journal of Multivariate Analysis*, 119:101–118.

Voev, V. (2008). Dynamic modelling of large-dimensional covariance matrices. In *High Frequency Financial Econometrics*, pages 293–312. Springer.

Whittaker, J. (2009). *Graphical models in applied multivariate statistics*. Wiley Publishing.

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. *Journal of Econometrics*, 160(1):33–47.

Zhang, L., Mykland, P. A., and Aït-Sahalia, Y. (2005). A tale of two time scales: Determining integrated volatility with noisy high-frequency data. *Journal of the American Statistical Association*, 100(472):1394–1411.

A Skewed generalized error distribution

The skewed generalized error distribution is specified by the location parameter μ, the scale parameter σ, the shape parameter ν and the skewness parameter ξ. Its density function is given by

$$f\left(\varepsilon|\mu,\sigma,\nu,\xi\right) = \frac{C}{\sigma}\exp\left(-\frac{|\varepsilon-\mu+\delta\sigma|^{\nu}}{\left[1-\operatorname{sign}\left(\varepsilon-\mu+\delta\sigma\right)\xi\right]^{\nu}\theta^{\nu}\sigma^{\nu}}\right)$$

with

$$C = \frac{\nu}{2\theta}\Gamma\left(\frac{1}{\nu}\right)^{-1},$$

$$\theta = \Gamma\left(\frac{1}{\nu}\right)^{1/2}\Gamma\left(\frac{3}{\nu}\right)^{-1/2}S\left(\xi\right)^{-1},$$

$$\delta = 2\xi A S\left(\xi\right)^{-1},$$

$$S\left(\xi\right) = \sqrt{1+3\xi^2-4A^2\xi^2},$$

$$A = \Gamma\left(\frac{2}{\nu}\right)\Gamma\left(\frac{1}{\nu}\right)^{-1/2}\Gamma\left(\frac{3}{\nu}\right)^{-1/2}.$$

For the parameter specification $\nu = 2$ and $\xi = 0$ the normal distribution is obtained.

B R-vine copula models

R-vine distributions are also referred to as pair-copula constructions, since they assign to each of the $d\left(d-1\right)/2$ edges of a d-dimensional R-vine structure a bivariate unconditional copula (in tree $\mathcal{T}_1$) or a bivariate conditional copula (in trees $\mathcal{T}_2$ to $\mathcal{T}_{d-1}$). We consider the copula data $(U_1,\ldots,U_d)$ corresponding to the random vector $(X_1,\ldots,X_d)$ with marginal distribution functions F_j $(j=1,\ldots,d)$, i.e. $U_j = F_j\left(X_j\right)$. Since in this case the marginals of the underlying data are uniform, we speak of an R-vine copula. Following Czado (2010), the d-dimensional R-vine copula density based on the R-vine structure $\mathcal{V}_d$ with edge set $E\left(\mathcal{V}_d\right) = E_1 \cup \cdots \cup E_{d-1}$ can be written as

$$\mathfrak{c}\left(u_1,\ldots,u_d\right) = \prod_{\ell=1}^{d-1}\prod_{e\in E_\ell}\mathfrak{c}_{a_e,b_e;D_e}\{\mathbb{C}_{a_e|D_e}\left(u_{a_e}|\boldsymbol{u}_{D_e}\right),\mathbb{C}_{b_e|D_e}\left(u_{b_e}|\boldsymbol{u}_{D_e}\right);\boldsymbol{u}_{D_e}\}, \qquad (9)$$

where

- $\mathfrak{c}_{a_e,b_e;D_e}\left(\cdot,\cdot;\boldsymbol{u}_{D_e}\right)$ denotes the copula density corresponding to the conditional distribution of $(U_{a_e},U_{b_e})'$ given $\boldsymbol{U}_{D_e} = \boldsymbol{u}_{D_e}$ with $\boldsymbol{U}_{D_e}$ the vector containing all variables corresponding to the conditioning set D_e. The corresponding copula will be denoted by $\mathbb{C}_{a_e,b_e;D_e}\left(\cdot,\cdot;\boldsymbol{u}_{D_e}\right)$.

- $\mathbb{C}_{a_e|D_e}\left(\cdot|\boldsymbol{u}_{D_e}\right)$ denotes the conditional distribution of U_{a_e} given $\boldsymbol{U}_{D_e} = \boldsymbol{u}_{D_e}$.

Given the large number of valid R-vine structures and given that the pair-copulas corresponding to each edge of the underlying R-vine structure can be chosen and combined arbitrarily R-vine copulas clearly constitute a highly flexible class of dependence models.

We assume that in (9) the conditional pair-copulas $\mathfrak{c}_{a_e,b_e;D_e}$ in trees $\mathcal{T}_\ell$ $(\ell = 2,\ldots,d-1)$ do not depend on the conditioning vector $\boldsymbol{u}_{D_e}$. Their arguments $\mathbb{C}_{a_e|D_e}\left(u_{a_e}|\boldsymbol{u}_{D_e}\right)$ and $\mathbb{C}_{b_e|D_e}\left(u_{b_e}|\boldsymbol{u}_{D_e}\right)$ indeed do depend on $\boldsymbol{u}_{D_e}$. For details on this simplifying assumption, see e.g. Hobæk Haff et al. (2010) and Stoeber et al. (2013).

Joe (1997) provides the important result for pair-copula constructions that the conditional distributions $\mathbb{C}_{a_e|D_e}\left(\cdot|\boldsymbol{u}_{D_e}\right)$ and $\mathbb{C}_{b_e|D_e}\left(\cdot|\boldsymbol{u}_{D_e}\right)$, subsequently abbreviated as $\mathbb{C}_{a|D}\left(\cdot|\boldsymbol{u}_D\right)$ and

$\mathbb{C}_{b|D}(\cdot|\boldsymbol{u}_D)$, can be evaluated using only the pair-copulas specified in lower tree levels of the underlying R-vine structure. Define for $i \in \{a, b\}$ the set $D_{+i} := D \cup \{i\}$. Then,

$$\mathbb{C}_{a|D_{+b}}(u_a|\boldsymbol{u}_{D_{+b}}) = h_{a|b;D}\{\mathbb{C}_{a|D}(u_a|\boldsymbol{u}_D)|\mathbb{C}_{b|D}(u_b|\boldsymbol{u}_D)\}$$

and

$$\mathbb{C}_{b|D_{+a}}(u_b|\boldsymbol{u}_{D_{+a}}) = h_{b|a;D}\{\mathbb{C}_{b|D}(u_b|\boldsymbol{u}_D)|\mathbb{C}_{a|D}(u_a|\boldsymbol{u}_D)\},$$

where

$$h_{a|b;D}\{\mathbb{C}_{a|D}(u_a|\boldsymbol{u}_D)|\mathbb{C}_{b|D}(u_b|\boldsymbol{u}_D)\} := \frac{\partial}{\partial u}\mathbb{C}_{a,b;D}\{\mathbb{C}_{a|D}(u_a|\boldsymbol{u}_D), u\}\bigg|_{u=\mathbb{C}_{b|D}(u_b|\boldsymbol{u}_D)}.$$

and

$$h_{b|a;D}\{\mathbb{C}_{b|D}(u_b|\boldsymbol{u}_D)|\mathbb{C}_{a|D}(u_a|\boldsymbol{u}_D)\} := \frac{\partial}{\partial u}\mathbb{C}_{a,b;D}\{u, \mathbb{C}_{b|D}(u_b|\boldsymbol{u}_D)\}\bigg|_{u=\mathbb{C}_{a|D}(u_a|\boldsymbol{u}_D)}.$$

are the h-functions corresponding to the pair-copula $\mathbb{C}_{a,b;D}$. Clearly, the arguments of the h-functions can again be expressed in terms of h-functions such that a recursive representation of $\mathbb{C}_{a|D}(u_a|\boldsymbol{u}_D)$ and $\mathbb{C}_{b|D}(u_b|\boldsymbol{u}_D)$ in terms of lower tree pair-copulas is obtained.

R-vine copulas have been extensively studied in the recent years including the development of comprehensive statistical software available in the R-package `VineCopula` (Schepsmeier et al. 2017).

C   Additional results for the empirical study

Table 5: RMSE with respect to the complete out-of-sample forecasting horizon (1632 days) for the model components in the Cholesky decomposition based model. The set of superior models according to the MCS approach at a confidence level of 10% is highlighted in gray. The lowest RMSE is highlighted in bold.

|	mean	HAR	HN	HSGED	ARFIMA	AN	ASGED
AXP,AXP	0.5215	0.2355	0.2358	0.2359	**0.2334**	0.2336	0.2340
AXP,C	0.6789	0.3730	0.3706	0.3769	0.3698	**0.3684**	0.3750
C,C	0.4579	0.2069	0.2076	0.2094	**0.2063**	0.2070	0.2081
AXP,GE	0.4142	0.2648	0.2646	0.2682	0.2618	**0.2616**	0.2659
C,GE	0.2431	0.1725	0.1726	0.1735	**0.1714**	0.1718	0.1727
GE,GE	0.3676	0.2111	0.2108	0.2112	0.2102	0.2099	**0.2096**
AXP,HD	0.4624	0.2976	0.2998	0.3023	**0.2944**	0.2972	0.2997
C,HD	0.2732	0.2160	0.2176	0.2165	0.2155	0.2163	**0.2153**
GE,HD	0.2352	0.1906	0.1913	0.1925	**0.1900**	0.1904	0.1917
HD,HD	0.3516	0.2165	0.2165	0.2169	0.2158	0.2157	**0.2156**
AXP,IBM	0.3102	0.2190	0.2180	0.2202	0.2174	**0.2166**	0.2189
C,IBM	0.1929	0.1529	0.1539	0.1534	**0.1527**	0.1533	0.1532
GE,IBM	0.1768	0.1401	0.1404	0.1409	**0.1399**	0.1401	0.1404
HD,IBM	0.1321	0.1260	0.1257	0.1254	0.1248	0.1246	**0.1246**
IBM,IBM	0.3390	0.2021	**0.2020**	0.2021	**0.2020**	0.2022	0.2026
AXP,JPM	0.6646	0.3918	0.3895	0.3946	0.3894	**0.3879**	0.3931
C,JPM	0.4075	0.2777	0.2804	0.2803	**0.2767**	0.2789	0.2784
GE,JPM	0.1981	0.1758	0.1757	0.1756	0.1745	**0.1743**	0.1744
HD,JPM	0.1619	0.1552	0.1554	0.1556	**0.1550**	0.1551	0.1556
IBM,JPM	0.1559	0.1498	0.1497	0.1500	**0.1496**	0.1498	0.1500
JPM,JPM	0.4461	0.2114	0.2112	0.2123	**0.2107**	**0.2107**	0.2116
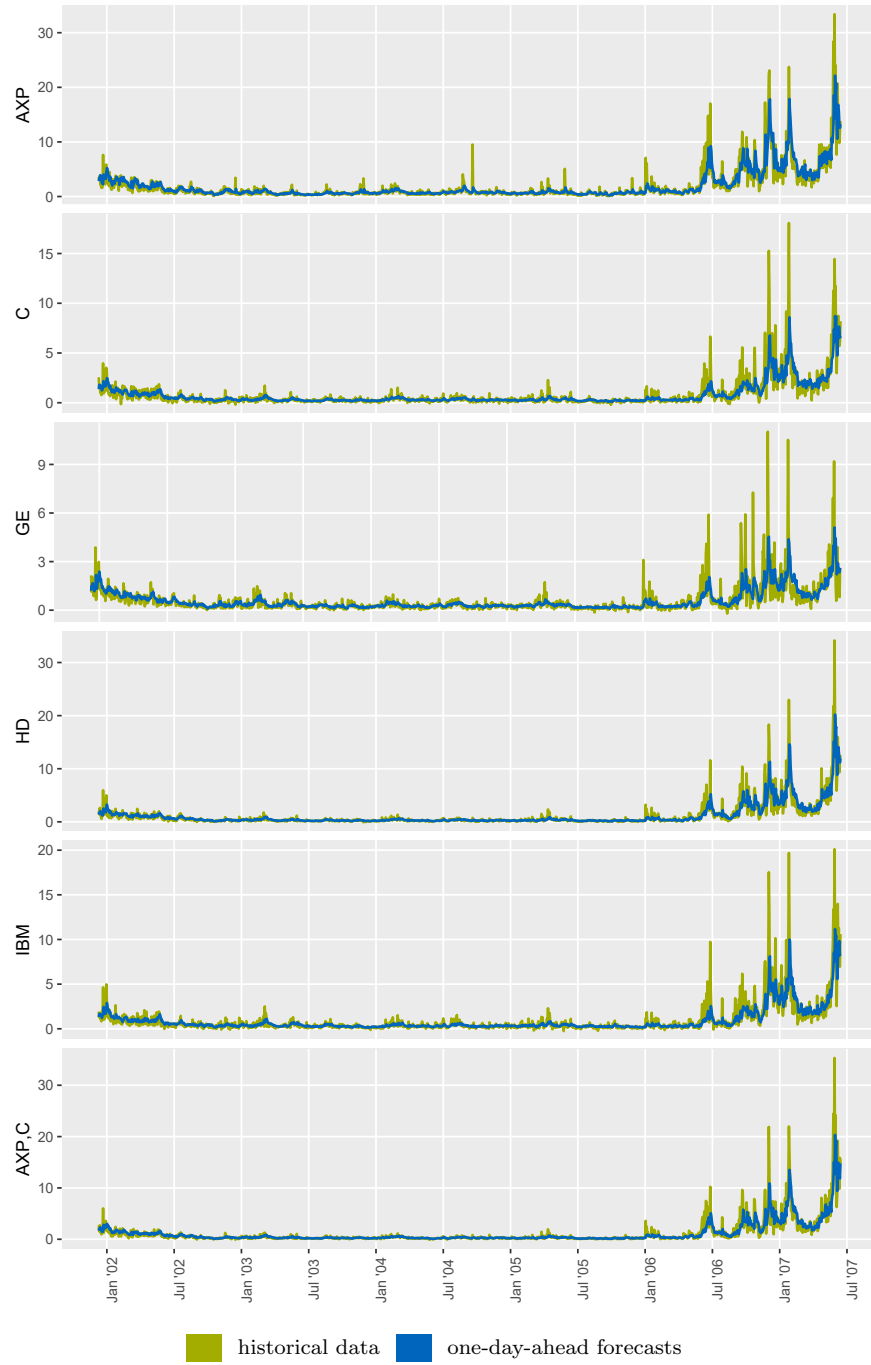
Figure 14: (part 1/3) Daily realized variance time-series and daily realized covariance time-series together with the time-series of the corresponding daily forecasts based on the partial correlation vine data transformation approach.
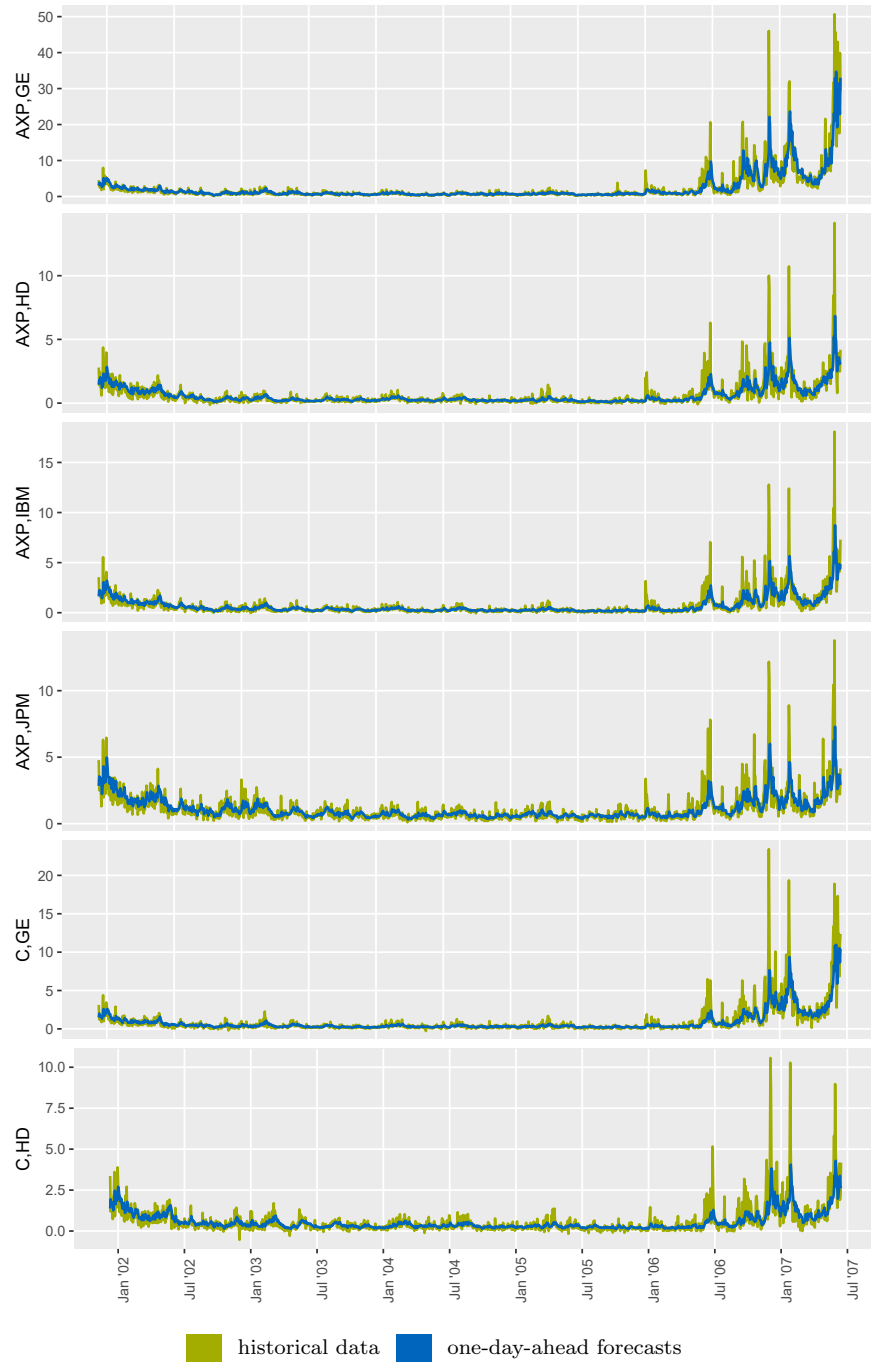
Figure 14: (part 2/3) Daily realized variance time-series and daily realized covariance time-series together with the time-series of the corresponding daily forecasts based on the partial correlation vine data transformation approach.
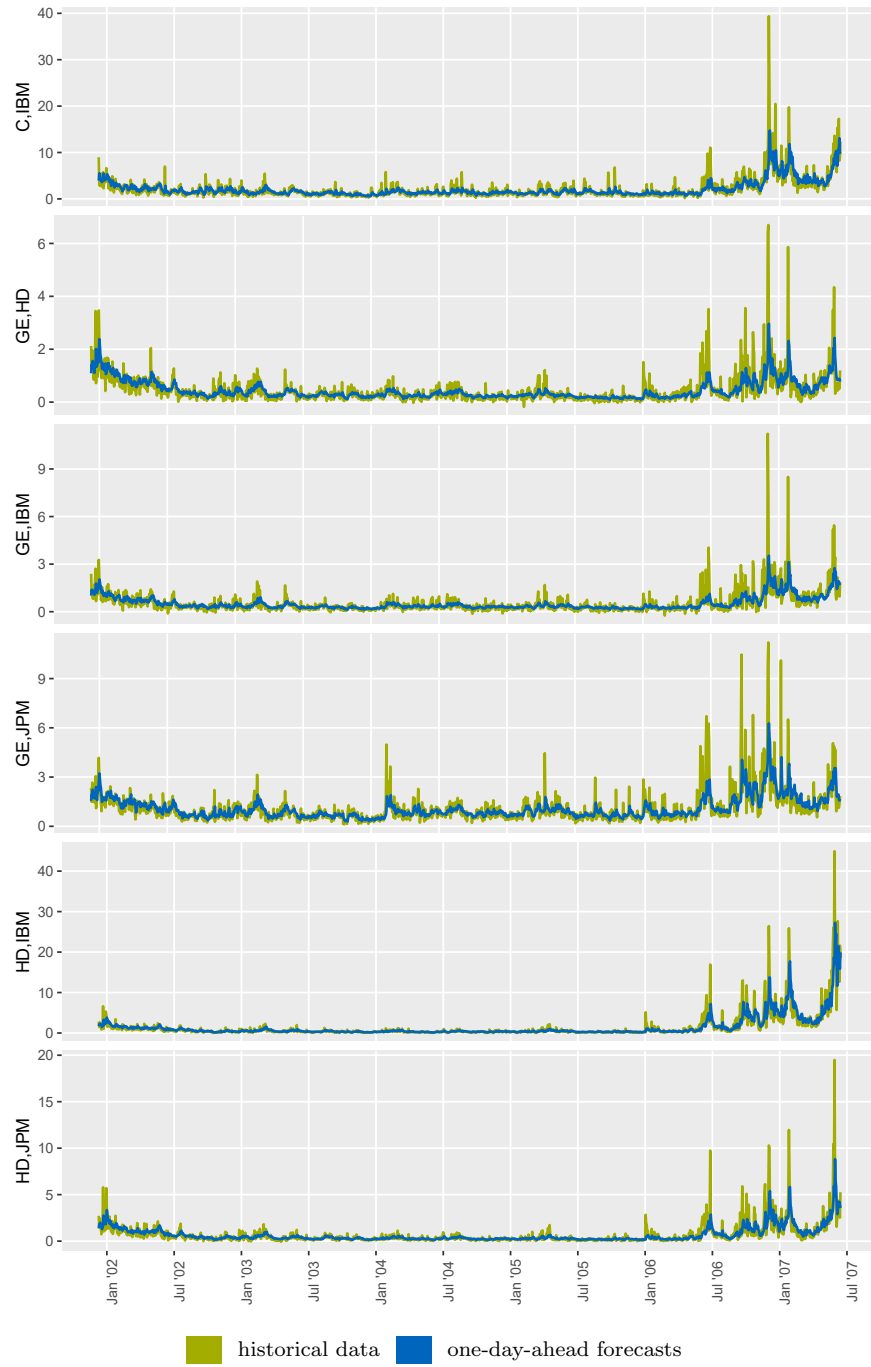
Figure 14: (part 3/3) Daily realized variance time-series and daily realized covariance time-series together with the time-series of the corresponding daily forecasts based on the partial correlation vine data transformation approach.