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

Nicole Barthel∗, Claudia Czado∗ and Yarema Okhrin†

February 28, 2018

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. A method to select a vine structure, which allows for parsimonious time-series modeling, is introduced. The predicted partial correlations have a clear practical interpretation. Being algebraically independent they do not underlie any algebraic constraint. The forecasting performance is evaluated through investigation of six-dimensional real data and is compared to Cholesky decomposition based benchmark models.

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 risks.

High-frequency data allow us 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 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 positive semi-definiteness and symmetry 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). Besides 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), data transformation is one of the most frequently used approaches to overcome these drawbacks. 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 semi-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 ARFIMA process. Brechmann et al. (2015) build upon this model approach, but pay special attention to the specific dependencies among the Cholesky series induced by the non-linear data transformation. While Cholesky decomposition based models are very straightforward and easy to implement, they also come with drawbacks. There is no clear interpretation of the parameterization and the latter induces an additive bias in the forecasts of the original data due to its non-linear nature. Also, the Cholesky decomposition depends on the ordering of the data, i.e. the ordering of the assets, with no obvious way to fix the order in advance. Complete enumeration as the only solution leads to a computationally extensive estimation problem. On the other hand, with a fixed ordering of the data, the Cholesky decomposition is static and does not allow dynamic adjustments to the underlying data over time. As an alternative, 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. Further, Andersen et al. (2006) develop 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.

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 fractionally integrated ARMA (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 parameterization tool for the realized covariance matrices, which allows for reasonable computational effort, interpretability of the decomposition and in particular to exploit the beneficial features of univariate time-series modeling, is desirable. A promising candidate which meets these requirements are partial correlation vines. They assign partial correlations to the edges of a so-called regular vine (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. Hence, out of the set of all partial correlations, which can be recursively calculated from standard correlations (Yule and Kendall 1965; Whittaker 2009), a partial correlation vine selects a subset of standard and partial correlations with attractive properties. First, Bedford and Cooke (2002) proof that there is a bijection between the (partial) correlations specified
by any R-vine structure and the set of symmetric and positive definite correlation matrices. Second, 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 associated correlation matrix. This result advocates partial correlation vines to be a useful tool in several applications. For example, 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 a 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 a 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, partial correlation vines as a parametrization tool of realized covariance matrices have not yet been investigated. We propose the so-called partial correlation vine approach, a joint prediction model of the realized variances and a subset of realized (partial) correlations specified by an R-vine structure. Each element of the latter subset corresponds to an edge of the underlying R-vine structure such that univariate series of realized (partial) correlations are obtained. This fact is of particular interest when developing a method to select among the large number of possible R-vine structures a parameterization best capturing the specific features of the underlying data. The proposed selection method exclusively relies on historical information contained in the modeled time-series and thus allows for a flexible and dynamic parameterization over time. Further, there is a clear practical and probabilistic interpretation leading to parsimonious time-series modeling in contrast to Cholesky decomposition based competitor models.

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. In Section 3, we give a detailed step-by-step instruction of the proposed partial correlation vine approach along with detailed discussion on R-vine structure selection in Section 3.2 and adequate time-series modeling in Section 3.3. 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 4 detailed investigation of the real data example will be continued. Studying the modeling and forecasting performance of the partial correlation vine approach compared to a Cholesky decomposition based benchmark model gives evidence for its usefulness and adequacy. This paper comes with extensive supplementary material.

2 Partial correlation vines

We present a dynamic framework for modeling and forecasting realized covariance matrices using partial correlation vines. First, we provide necessary background on the two main ingredients of the proposed parameterization tool – partial correlations and regular vines.

2.1 Partial correlations

We consider a random vector \( X_T := (X_1, \ldots, X_d), d \geq 2, \) with zero mean, where \( \mathcal{I} \) is the index set \( \{1, \ldots, d\} \). We denote the \( d \times d \) covariance matrix by \( Y \) and obtain the associated \( d \times d \) correlation matrix \( R \) through the relationship \( R = D^{-1/2} Y D^{-1/2} \), where \( D = \text{diag}(y_{1,1}, \ldots, y_{d,d}) \) 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 \(D_{\{i,j\}} := L_{\sim \{i,j\}} = L \setminus \{i,j\}\) and the associated random vector by \(X_{D_{\{i,j\}}} := \{X_k, k \in D_{\{i,j\}}\}\). The partial regression coefficients \(b_{i,j:D_{\{i,j\}}}\) are defined as the quantities that minimize

\[
E[(X_i - \sum_{j \in L_{\sim \{i\}}} b_{i,j:D_{\{i,j\}}} X_j)^2].
\]

The associated partial correlation coefficients \(\rho_{i,j:D_{\{i,j\}}}\) quantify the dependence between \(X_i\) and \(X_j\) without the linear effect of \(X_D\) and are defined by

\[
\rho_{i,j:D_{\{i,j\}}} := \text{sgn}(b_{i,j:D_{\{i,j\}}}) \left( b_{i,j:D_{\{i,j\}}} b_{j,i:D_{\{i,j\}}} \right)^{1/2}.
\]

We refer to the cardinality of \(D_{\{i,j\}}\) as order of the associated 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 \(\tilde{D} := L_{\sim \{i,j,k\}}\) such that \(D_{\{i,j\}} = \tilde{D} \cup k\). Yule and Kendall (1965) give a formula to calculate the partial correlations of order \(|D_{\{i,j\}}|\) with \(|D_{\{i,j\}}| \geq 1\) in terms of (partial) correlations of lower order. With \(\rho_{i,k;\tilde{D}}^2 < 1\) and \(\rho_{j,k;\tilde{D}}^2 < 1\) it holds that

\[
\rho_{i,j:D_{\{i,j\}}} = \frac{\rho_{i,j;\tilde{D}} - \rho_{i,k;\tilde{D}} \rho_{j,k;\tilde{D}}}{\sqrt{1 - \rho_{i,k;\tilde{D}}^2} \sqrt{1 - \rho_{j,k;\tilde{D}}^2}}.
\]

Thus, for the pair \((i, j), i, j \in \mathcal{I}, i \neq j\), partial correlations of any order \(p, 1 \leq p \leq d - 2\), can be calculated from standard correlations through recursive application of (1).

In the following, we refer to \(\mathcal{C}_d\) as the set of all standard correlations and to \(\mathcal{C}^p_d\) as the set of all pairwise standard and partial correlations. Further, we introduce the vector \(\mathbf{P}_d\), resp. \(\mathbf{P}^p_d\), which collects for the random vector \(X_T\) the values of the associated \((\binom{d}{2})\) standard correlations, resp. the values of the \((\binom{d}{2}) 2^{d-2}\) standard and partial correlation coefficients. We use a lexicographical ordering of the elements with increasing subset \(L \subseteq \mathcal{I}\), i.e.

\[
\mathbf{P}_d := \left( \rho_{1,2}, \rho_{1,3}, \ldots, \rho_{1,d}, \rho_{2,3}, \ldots, \rho_{2,d}, \ldots, \rho_{(d-1),d} \right)
\]

and

\[
\mathbf{P}^p_d := \left( \mathbf{P}_d, \rho_{1,2;3}, \ldots, \rho_{1,2;d}, \rho_{1,d;2}, \ldots, \rho_{1,d;\{d-1\}}, \rho_{2,3;1}, \ldots, \rho_{(d-1),d;\{d-2\}}, \ldots, \rho_{1,2;3,\ldots,d}, \rho_{1,3;2,4,\ldots,d}, \ldots, \rho_{1,d;2,\ldots,\{d-1\}}, \ldots, \rho_{(d-1),d;1,\ldots,(d-2)} \right).
\]

To sum up, as shown in Figure 1 we retain from a \(d \times d\) variance-covariance matrix \(\mathbf{Y}\) the \(d\)-dimensional vector of variances \(\mathbf{y}\) together with a \((\binom{d}{2}) 2^{d-2}\)-dimensional vector \(\mathbf{P}^p_d\), which takes values in \((-1,1)\) and collects all standard and partial correlations associated to \(\mathbf{Y}\).

### 2.2 R-vines

Having the set of all standard and partial correlations \(\mathcal{C}^p_d\) at hand, we introduce regular vines (R-vines) as a tool to select a subset out of \(\mathcal{C}^p_d\), which exhibits attractive properties when interest is in dynamic modeling and forecasting of realized correlation matrices.

According to Bedford and Cooke (2002), an R-vine on \(d\) elements is a set of \(d - 1\) linked trees, i.e. undirected and acyclic graphs, \(V_d := (T_1, \ldots, T_{d-1})\) with the set of edges \(E(V_d) := E_1 \cup \cdots \cup E_{d-1}\) and the set of nodes \(N(V_d) := N_1 \cup \cdots \cup N_{d-1}\) such that

(i) \(T_1\) is a tree with nodes \(N_1 = \{1, \ldots, d\}\) and edges \(E_1\),

(ii) ...
Data: random vector $X_I \in \mathbb{R}^d$ with covariance matrix $Y \in \mathbb{R}^{d \times d}$

variance vector of $X_I$

$y = (y_1, \ldots, y_d) \in \mathbb{R}^d$

correlation matrix of $X_I$

$R = \begin{pmatrix} 1 & \rho_{1,2} & \cdots & \rho_{1,d} \\ \rho_{1,2} & 1 & \cdots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{1,d} & \cdots & \cdots & 1 \end{pmatrix} \in (-1,1)^{d \times d}$

vector of standard correlations of $X_I$

$P_{c_d} = (\rho_{1,2}, \ldots, \rho_{d-1,d}) \in (-1,1)^{\binom{d}{2}}$

vector of standard and partial correlations of $X_I$

$P_{c_d}^{\ast} = (\rho_{1,2}, \ldots, \rho_{d-1,d}, \rho_{1,2}, \ldots, \rho_{d-2,d}) \in (-1,1)^{\binom{d}{2} + d^2 - 2}$

Figure 1: Data retained from a given covariance matrix $Y$.

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

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

According to property (ii), the $d - (\ell - 1)$ edges $E_{\ell-1}$ in $T_{\ell-1}$ become nodes in $T_\ell$. 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 $\mathcal{V}_d$, $2 \leq \ell \leq d - 1$, with $a, b \in N_\ell$. Its complete union $U_e^\ast$ is the subset of nodes in $T_1$, i.e. the subset of $\{1, \ldots, d\}$, reachable from $e$ by the membership relation, i.e.

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

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

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

The associated symmetric difference is referred to as conditioned set

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

By definition, each conditioned set in $\mathcal{V}_d$ consists of two single elements and in particular forms a unique pair of variables $i, j \in \{1, \ldots, d\}, i \neq j$.

Example 2.1. Figure 2 shows an R-vine structure on six elements labeled with the conditioned set and the conditioning set associated to each edge. The latter is indicated by a leading "\$\$". The bold tree segment in $T_2$ corresponds to the edge $e = \{\{1, 2\}, \{2, 6\}\}$. Reachable from edge $\{1, 2\} \in T_1$, resp. $\{2, 6\} \in T_1$, are the nodes $1, 2 \in N_1$, resp. $2, 6 \in N_1$. Thus, $D_e = U_{\{1,2\}}^\ast \cap U_{\{2,6\}}^\ast = \{1, 2\} \cap \{2, 6\} = \{2\}$ is the conditioning set associated to $e$ and the conditioned set is $\{C_{e,\{1,2\}}, C_{e,\{2,6\}}\} = \{\{1, 2\} \setminus \{2\}, \{2, 6\} \setminus \{2\}\} = \{1, 6\}$. 

5
Having the definitions for both the partial correlations associated to a realized correlation matrix and the specification of an R-vine structure \( V_d \) at hand, building the bridge between these two ingredients is straightforward: In a partial correlation vine with R-vine structure \( V_d \) each edge \( e = \{a, b\} \in E(V_d) \) is identified with the partial correlation coefficient, which coincides with the conditional bivariate constraint specified by \( e \), i.e. with \( \rho_{C_{e,a},C_{e,b};D_e} \). By doing so, to each edge a value in \((-1, 1)\) is assigned. We define the set of the \( \binom{d}{2} \) standard and partial correlations specified by \( V_d \) as

\[
C(V_d) := \{ \rho_{C_{e,a},C_{e,b};D_e} : e = \{a, b\} \in E(V_d), \rho_{C_{e,a},C_{e,b};D_e} \in \mathcal{C}^p_d \}.
\]

The vector \( P_C(V_d) \in (-1, 1)^{\binom{d}{2}} \) collects for \( X_T \) the associated values in lexicographical order. 

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

\[
F_{PCor2Cor} : (-1, 1)^{\binom{d}{2}} \rightarrow (-1, 1)^{\binom{d}{2}}, \quad F_{PCor2Cor}(P_C(V_d)) = P_{C_d}.
\] (2)

As a direct consequence, we can deduce \( C_d \) from any set \( C(V_d) \). The calculation procedure is based on recursion formula (1) and is further discussed in Lewandowski et al. (2009). An efficient implementation of the bijection \( F_{PCor2Cor} \) is provided in the R-package VineCopula with the function RVineCor2Cor (Schepsmeier et al. 2017).

**Example 2.1** (continued). The R-vine structure on 6 elements illustrated in Figure 2 specifies the partial correlations given in Table 1. As an example, we calculate the standard correlation \( \rho_{3,6} \) from the partial correlations highlighted in bold in Table 1. We start from the unique edge, for which \( \{3,6\} \) forms the conditioned set, i.e. in \( T_3 \). As conditioning set, we have \( \{1,2\} \). Since 1 does not appear in the conditioning sets of lower tree levels, we exclude 1. Based on (1) we have

\[
\rho_{3,6;1,2} = \frac{\rho_{3,6,2} - \rho_{1,3,2}\rho_{1,6,2}}{\sqrt{1 - \rho_{1,3,2}^2}\sqrt{1 - \rho_{1,6,2}^2}}
\]

with all values on the right-hand side known except for \( \rho_{3,6,2} \). The latter is obtained via

\[
\rho_{3,6,2} = \rho_{3,6;1,2}\sqrt{1 - \rho_{1,3,2}^2}\sqrt{1 - \rho_{1,6,2}^2} + \rho_{1,3,2}\rho_{1,6,2}.
\]
In a similar way, we proceed in the next lower tree level, which already is \( T_1 \) specifying standard correlations. Based on (1), we can write
\[
\rho_{3,6;2} = \frac{\rho_{3,6} - \rho_{2,3} \rho_{2,6}}{\sqrt{1 - \rho_{2,3}^2} \sqrt{1 - \rho_{2,6}^2}}
\]
and finally we obtain
\[
\rho_{3,6} = \rho_{3,6;2} \sqrt{1 - \rho_{2,3}^2} \sqrt{1 - \rho_{2,6}^2} + \rho_{2,3} \rho_{2,6}.
\]

Table 1: Partial correlations specified by the R-vine structure given in Figure 2.

| tree level | partial correlation order | #partial correlations | partial correlations |
|------------|---------------------------|-----------------------|---------------------|
| \( T_1 \)  | 0                         | 5 (= d - 1)           | \( \rho_{2,6}, \rho_{1,2}, \rho_{2,4}, \rho_{2,3}, \rho_{3,5} \) |
| \( T_2 \)  | 1                         | 4 (= d - 2)           | \( \rho_{1,6}, \rho_{1,3}, \rho_{3,4}, \rho_{2,5,3} \) |
| \( T_3 \)  | 2                         | 3 (= d - 3)           | \( \rho_{3,6}, \rho_{1,2}, \rho_{4,2,3}, \rho_{4,5,2,3} \) |
| \( T_4 \)  | 3                         | 2 (= d - 4)           | \( \rho_{4,6}, \rho_{1,2,3}, \rho_{3,4,2,3} \) |
| \( T_5 \)  | 4                         | 1 (= d - 5)           | \( \rho_{5,6,1,2,3,4} \) |

To the theory on partial correlation vines Kurowicka and Cooke (2006) add the essential result that any correlation matrix (built from the values in \( P_{C_d} \)) derived from a partial correlation vine (consisting of the values in \( P_{C(V_d)} \)), is positive definite with correlation values in \((-1,1)\) for all off-diagonal elements. Note that this is true for any arbitrary assignment of values in \((-1,1)\) to the edges of an R-vine structure. Thus, the elements in \( C(V_d) \) are algebraically independent, which is the key result for the proposed vine based modeling approach: By modeling and forecasting the time-series of partial correlation vines, we obtain forecasts, which do not underly any algebraic restrictions but on the contrary automatically guarantee positive definiteness of the associated predicted correlation matrices. Figure 3 summarizes the introduced relationships between the sets \( C_d, C_d^p \) and \( C(V_d) \) and illustrates the parameterization of \( C_d \) through a partial correlation vine.

![Figure 3](image-url)

Figure 3: Illustration of the parameterization of the set of standard correlations \( C_d \) through a partial correlation vine, which selects a subset of algebraic independent (partial) correlations \( C(V_d) \subset C_d^p \) from all standard and partial correlations. The abbreviation “pcor” is used for partial correlation.
3 Partial correlation vine approach

In this section, we give a step-by-step guidance through the proposed partial correlation vine based model for prediction of realized covariance matrices. Throughout, the proceeding is supported by a real data example: We consider high frequency data from the NYSE TAQ database containing tick-by-tick bid and ask quotes on six stocks listed on the NYSE, American Stock Exchange (AMEX), and the National Association of Security Dealers Automated Quotation System (NASDAQ). Data of 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. The original raw data was processed by Chiriac and Voev (2011) to obtain the series of 2156 realized covariance matrices.

3.1 General setting and data characteristics

Based on the daily price series $S_t \in \mathbb{R}^d$, $t = 1, \ldots, T$, for $d$ assets, interest is in the vector of log-returns $r_t = \log(S_t) - \log(S_{t-1})$ with $d$-dimensional conditional mean vector $\mathbb{E}[r_t | F_{t-1}]$ and $d \times d$-dimensional conditional covariance matrix $\Sigma_t$. Here, $F_{t-1}$ is the information set containing all information up to and including time point $t - 1$. Using high-frequency data, realized covariance matrices can be obtained as consistent nonparametric estimates for $\Sigma_t$ (Barndorff-Nielsen and Shephard 2004). For each day $t$, we consider $M$ intra-day periods with price series $S_{\ell,t} \in \mathbb{R}^d$, $\ell = 1, \ldots, M$, and vectors of log-returns $r_{\ell,t} = \log(S_{\ell-1+t}/M) - \log(S_{t-1+(\ell-1)/M})$. The realized covariance matrix is defined as

$$Y_t = \sum_{\ell=1}^M r_{\ell,t} r_{\ell,t}', \ t = 1, \ldots, T.$$

By calculating realized variances and covariances, naturally latent variables, namely daily integrated volatility and covariance, become measurable and observable. From the daily series of realized covariance matrices $Y_t$, $t = 1, \ldots, T$, the associated daily time-series for realized variances $y_t = (y_{1,1,t}, \ldots, y_{d,d,t})$ and realized correlation matrices $R_t$ can be derived. Using recursion (1), for each day $t$ the vector of standard and partial correlations $P_{c_t^R} \in (-1,1)^{d^2/2}$ is obtained (Figure 1). To scale the data to values in $\mathbb{R}$ we log-transform the variances and apply the Fisher $z$-transformation to the elements $\rho_{(j):t}$ of $P_{c_t^R}$ with $(j) \in \{(1,2), \ldots, (d-1,d), (1,2;3), \ldots, (d-1,d;1,\ldots,d-2)\}$, i.e.

$$z(\rho_{(j):t}) = \frac{1}{2} \log \left( \frac{1 + \rho_{(j):t}}{1 - \rho_{(j):t}} \right). \quad (3)$$

Figure 4 shows for the considered real life data a selection of time-series. The first panel illustrates for JPM the daily realized variance series both on the original and on the log-transformed scale. Striking is the highly volatile behavior particularly during periods of financial turmoil such as the aftermaths of the dotcom bubble and the beginning of the financial crisis in August 2007. Panels 2 to 6 show selected daily time-series of (partial) correlations with increasing order, e.g. the correlation time-series for Citigroup and JPM is illustrated in the second row; the last panel illustrates the time-series of dependence between IBM and JPM with the effect of all four remaining stocks removed. With increasing order the partial correlation time-series become more stable while still exhibiting highly volatile behavior with values significantly larger than zero. Given that in total there are $240$ ($d = 6$) (partial) correlation time-series, the natural question is how to decide on a specific R-vine structure $\mathcal{V}_6$ to select a subset $\mathcal{C}(\mathcal{V}_6) \subset C_6^R$ of 15 (partial) correlations for the data at hand.
**Figure 4:** Daily realized variance series (1st row) and daily realized (partial) correlation series (2nd – 6th row). Original data is shown on the left, log-transformed data, resp. Fisher z-transformed data, is shown on the right.
Table 2: Illustration of the R-vine structure selection method for the real data example considering all available data points, i.e. the mean values $\bar{\rho}_{c_{e,a},c_{e,b},d_{e}}^{t}$ are based on $t = 1, \ldots, 2156$.

| $\mathcal{D}_{e}$ | $c_{e,a},c_{e,b}$ | $\bar{\rho}_{c_{e,a},c_{e,b},d_{e}}^{t}$ | selected tree |
|-------------------|-------------------|---------------------------|---------------|
| 0                 | C,JPM             | 0.547                     |               |
| 0                 | AXP,C             | 0.456                     |               |
| 0                 | C,GE              | 0.437                     |               |
| 0                 | AXP,JPM           | 0.433                     |               |
| 0                 | GE,IBM            | 0.400                     |               |
| 0                 | AXP,GE            | 0.394                     |               |
| 0                 | GE,JPM            | 0.393                     |               |
| 0                 | AXP,IBM           | 0.362                     |               |
| 0                 | AXPC              | 0.358                     |               |
| 0                 | C,HD              | 0.355                     |               |
| 0                 | GE,HD             | 0.352                     |               |
| 0                 | AXP,HD            | 0.333                     |               |
| 0                 | HD,JPM            | 0.333                     |               |
| 0                 | HD,IBM            | 0.330                     |               |
| GE                | C,IBM             | 0.253                     |               |
| C                 | AXP,JPM           | 0.247                     |               |
| C                 | AXP,GE            | 0.241                     |               |
| C                 | GE,HD             | 0.229                     |               |
| 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                     |               |
| C,GE              | AXP,HD            | 0.154                     |               |
| C,GE,IBM          | AXP,HD            | 0.129                     |               |
| AXP,C,GE          | IBM,JPM           | 0.121                     |               |
| AXP,C,GE,IBM      | HD,JPM            | 0.093                     |               |
3.2 R-vine structure selection

In $d$ dimensions there exist $d!/2 \cdot d^{(d-2)(d-3)/2}$ valid R-vine structures (Morales Napoles et al. 2010) and thus, in theory, as many selectable subsets $C(V_d) \subset C_d^P$. As will be shown, by careful selection of a specific R-vine structure parsimonious time-series modeling can be achieved.

We use that for a series of partial correlation vines with same R-vine structure, each edge is identified with a univariate time-series of standard or partial correlation through the associated bivariate constraint, i.e. each edge can be characterized by a weight derived from sample properties of the associated time-series. We decide for the average (partial) correlation strengths: we consider the average correlation matrix $\bar{R} = (\bar{\rho}_{i,j})_{i,j=1}^d$ (which is positive definite) calculated from $R_t$, $t = 1, \ldots, T$, and obtain all associated (partial) correlations $P_{ij}$ through recursion formula (1). Then, within a top-down procedure we find tree by tree ($\ell = 1, \ldots, d-1$) the maximum spanning tree $T_\ell$ (Katoh et al. 1981) with edge weights set to $\bar{\rho}_{C_e,a;C_e,b;D_e}$ for $|D_e| = \ell - 1$ and $(C_e,a;C_e,b;D_e)$ satisfies the proximity condition given $T_1$ to $T_{\ell-1}$. By doing so, we equip based on historical information the partial correlation vine $V_d$ with the on average strongest (partial) correlation means. Based on the so-obtained subset $C(V_d) \subset C_d^P$ the vectors $P_{C(V_d),t}$, $t = 1, \ldots, T$, can be constructed.

The correlation matrix $\bar{R}$ can be obtained in various ways depending on how the average is calculated. Considering for each pair $i, j \in \{1, \ldots, d\}$, $i \neq j$, 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. 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 $\lambda$ the impact of more recent days increases and therewith the sensitivity of the parameterization to market changes. This approach will be utilized in the data application.

For the data at hand, the proposed R-vine structure selection method is illustrated in Table 2. As edge weights the empirical means of the (partial) correlation series based on all data points, i.e. $t = 1, \ldots, 2156$ (January 1, 2000 - July 30, 2008), are chosen. In $T_1$, 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 $T_2$ only edges satisfying the proximity condition given $T_1$ are allowed. A tree is constructed by the four edges with the highest mean values of 1st order partial correlations, etc.

The proposed R-vine structure selection method is motivated by the definition of partial correlation vines. Each edge has a unique conditioned set $i, j \in \{1, \ldots, d\}$, $i \neq j$, i.e. depending on the tree level $\ell$ the dependence between $X_i$ and $X_j$ is either captured through the standard correlation (if $(i,j)$ occurs as conditioned set in $T_1$) or through a $(\ell-1)$-th order partial correlation (if $(i,j)$ occurs as conditioned set in $T_\ell$, $\ell = 2, \ldots, d-1$). In the latter case the linear effect of the $(\ell-1)$ variables $X_{D_e}$, $D_e \subset \{1, \ldots, d\}\setminus\{i,j\}$, $|D_e| = \ell - 1$, on the dependence between $X_i$ and $X_j$ is removed. The implications of this construction principle become clear for the real data example. It includes three market participants of financial sectors, namely American Express Inc. (AXP), Citigroup (C) and JPMorgen Chase & Co (JPM), International Business Machines (IBM) as an IT service, Home Depot Inc. (HD) representing building materials trade and the diversified industrial corporation General Electric (GE). As can be seen in Table 2 the strong pairwise correlations between the three financial services are captured in trees $T_1$ and $T_2$. From $T_3$ on only partial correlations between stocks from different market sectors are modeled. In Figure 5 data characteristics of three selected time-series are illustrated. The figures in the top row rely on the realized variance series for JPM, which together with the remaining five variance series always needs to be modeled. From the long hyperbolic decay of the autocorrelation function of the squared data on the left long-memory behavior and the presence of volatility clustering can be detected. The log-periodogram shows higher peaks only for short frequencies as expected for self-similar processes. These are typical characteristics observed for financial volatility data, which require
elaborate and careful time-series modeling. In the second row, resp. third row, an exemplary time-series appearing in $T_1$, resp. $T_4$, of the selected R-vine structure is considered. Interestingly, while the standard correlation time-series in $T_1$ inherits the data characteristics of the variance time-series, the latter are less pronounced for the partial correlation time-series in $T_4$. Given the selection criterion of strong average correlation strengths, in higher trees only variable pairs are modeled, of which the standard correlation might mainly be driven by other variables. Once this influence is removed the resulting partial correlation time-series behave more and more like noise allowing for parsimonious time-series modeling. Thus, the proposed R-vine structure selection method results in an inhomogeneous data complexity of the time-series, which need to be modeled. Note that the partial correlation time-series underlying the two figures in the last row of Figure 5 shows similar data characteristics as the variance time-series in the top row even though it would occur in $T_5$ of an R-vine structure. However, given the proposed R-vine structure selection method it was not selected.

Figure 5: Illustration of data characteristics of the volatility time-series. In the first row, the log-transformed time-series of JPM is considered. In rows 2 to 4 exemplary Fisher-z transformed (partial) correlation time-series of increasing order are considered. On the left autocorrelation functions of squared data and on the right corresponding log-periodograms are shown. For all time-series the period from July 1, 2006 to June 30, 2008 is considered.
3.3 Time-series modeling and forecasting

Given the data characteristics detected in the previous section appropriate time-series models are needed for estimation of the $d$ log-transformed variances and the $(d^2)$ selected Fisher z-transformed (partial) correlation time-series (see (3)). This requires time-series modeling in $(d^2 + d)$ dimensions. Often, multivariate time-series models lack parameter flexibility and the ability to properly capture non-Gaussianity or heteroscedasticity. A copula based time-series model allows to use the flexibility of univariate time-series models to tackle these problems.

A $d$-dimensional copula $C$ is a distribution function on $[0,1]^d$ with uniformly distributed margins. Sklar (1959) provides the fundamental theorem, that the joint distribution function $F$ of a random vector $Z_I = (Z_1,\ldots,Z_d)$ can be expressed in terms of its marginal distributions $F_j, j = 1,\ldots,d$, and its associated copula, i.e. $F(z_I) = C(F_1(z_1),\ldots,F_d(z_d))$. The copula $C$ is unique if $Z_I$ is absolutely continuous. As a direct consequence of Sklar’s Theorem (Sklar 1959) marginal and dependence modeling for the joint distribution $F$ can be handled separately. First, estimates $\hat{F}_1,\ldots,\hat{F}_d$ are obtained. Then, by applying the probability integral transform, pseudo random variables on the so-called $u$-scale, which are approximately uniformly distributed on $[0,1]$, are obtained. We have $\hat{U}_j := \hat{F}_j(Z_j), j = 1,\ldots,d$, of which the joint distribution function is the copula $C$ associated to $F$. See Joe and Xu (1996) and Joe (2005) for a thorough investigation of this two-stage proceeding called inference for margins (IFM). Here, a sample of independent data $(z_1,\ldots,z_d)_{i=1}^n$ is needed. Thus, univariate time-series models need to be applied to the marginals first to remove serial dependence.

Marginal time-series modeling

Let $\eta_t$ be a log-transformed variance, resp. Fisher-z transformed (partial) correlation, at time $t$. We consider HAR models (Corsi 2009), which incorporate one day ($d$), one week ($w$) and one month ($m$) averages $\eta_{t-1}, \eta_{t-1}^{(5)}$ and $\eta_{t-1}^{(22)}$. The latter two are derived from the aggregated covariance matrix over the last $w$ and $m$ days. Including shorter and longer term volatilities, HAR models account for different time horizons and their interconnectedness with regard to financial decision making. This allows to capture data features like long persistence and multifractal scaling. With $w = 5$ and $m = 22$ the basic HAR model is given by

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

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

A possible extension to account for the inhomogeneous data complexity within the partial correlation vine approach is a GARCH (1,1) component to capture non-Gaussianity and volatility clustering. The innovations can either be assumed to be Gaussian or to follow e.g. a skewed generalized error distribution (SGED) (Bai et al. 2003; Corsi et al. 2008; Fernández and Steel 1998). The latter specification additionally captures possible high kurtosis and skewness of residuals (see Section A in the supplementary material for details).

Dependence modeling

Application of the above time-series models to the univariate marginals yields multivariate residuals. The latter inherit the cross-sectional dependencies within the original multivariate time-series data, but are no longer subject to the serial dependence inherent in the individual time-series. The parametric marginal distributions of the residuals are specified through the associated fitted time-series models, such that pseudo copula data can be calculated. To them an appropriate copula model can be fitted, e.g. via maximum likelihood optimization.
Forecasting

From the fitted copula model for each day of the prediction horizon out-of-sample forecasts on the u-scale can be sampled. The simulated data exhibit the dependence structure as estimated for the training set data. Applying the inverse probability integral transform gives predictions for the residuals. Based on the associated fits of the univariate time-series models predictions of the log-transformed variances and Fisher-z transformed (partial) correlations are obtained. Finally, using the bijection $F_{PCor2Cor}$ (see (2)) gives forecasts for the correlation matrices and when combined with the predicted variance forecasts for the realized covariances.

3.4 Modeling approach at a glance

To summarize the proposed modeling strategy Figure 6 provides a step-by-step guide.

**Data:** series of daily realized covariance matrices $Y_t, t = 1, \ldots, T$

| Series of realized variance vectors $y_t, t = 1, \ldots, T$ | Series of realized (partial) correlation vectors specified by $V_d$ $PC(V_d), t = 1, \ldots, T$ |
|---|---|
| log-transformation | Fisher z-transformation |
| series of data vectors scaled to $\mathbb{R}^{d(d+1)/2}$ $P_t = (p_1, \ldots, p_d(d+1)/2, t = 1, \ldots, T$ |

Section 3.3

model $P_t, t = 1, \ldots, T$, through a copula based approach including HAR based univariate time-series models

| Series of forecasts $\hat{P}_t = (\hat{p}_1, \ldots, \hat{p}_d(d+1)/2, t = T + 1, \ldots, T + \hat{T}$ |
|---|
| exp-transformation | inverse Fisher z-transformation |
| series of predicted variances $\hat{y}_t, t = T + 1, \ldots, T + \hat{T}$ | series of predicted partial correlation vines built from $\hat{P}_C(V_d), t = T + 1, \ldots, T + \hat{T}$ |
| | $F_{PCor2Cor}$, Section 2 |
| | series of predicted standard correlations $\hat{P}_{C_d}, t = T + 1, \ldots, T + \hat{T}$ |

series of predicted covariance matrices $\hat{Y}_t, t = T + 1, \ldots, T + \hat{T}$

**Figure 6:** Partial correlation vine approach for training set covering days $t = 1, \ldots, T$ and test set covering days $t = T + 1, \ldots, T + \hat{T}$. 

14
3.5 Cholesky decomposition based benchmark model

To investigate the forecasting potential and performance of the proposed partial correlation vine approach a Cholesky decomposition based benchmark model is considered. Relying on this frequently used modeling strategy e.g. Chiriac and Voev (2011) and Brechmann et al. (2015) investigate the real data example introduced in Section 3. The basic idea is to apply for each day \( t, t = 1, \ldots, T \), the Cholesky decomposition to the realized covariance matrices \( Y_t \), i.e. \( Y_t = P_t'P_t \), where \( P_t' \) is a lower triangular matrix with positive diagonal elements. Time-series models are applied to the Cholesky elements, i.e. the \( \binom{n}{2} + d \) entries of \( P_t' \). Positive definiteness of the predicted covariance matrices is automatically guaranteed as long as their predicted lower triangular Cholesky counterparts have positive diagonal entries. Note that the series of Cholesky elements obtained from the realized covariance matrices depends on the ordering of the assets within the matrices. For a fixed order on the other hand the Cholesky decomposition is unique and therewith static over time. For the six stocks in the real data example we assume alphabetical ordering. Application of the Cholesky decomposition to the series of realized covariance matrices yields time-series of 21 Cholesky elements.

4 Empirical study

The real data example introduced in Section 3 will now be investigated in more detail. Forecasts of the realized covariance matrices will be estimated and compared to those of the Cholesky decomposition based benchmark model with regard to robustness and accuracy.

4.1 Moving window approach

Data for the period from January 1, 2000 until June 30, 2008 are available, i.e. for 2156 days. For modeling and forecasting we utilize a moving window approach. 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 the considered HAR models rely on a monthly average of 22 days, the first forecast is obtained for day 525. In total, there are 75 time windows. Figure 7 illustrates the moving window approach.

![Figure 7: Moving window approach illustrated for the considered real data example.](image)

4.2 Dynamic parameterization

Since the ordering of the assets remains the same over time and the Cholesky decomposition is unique in that case, the same Cholesky elements are modeled for all time windows \( W_i, i = 1, \ldots, 75 \). However, there is a large number of possible R-vine structures. Thus, in the partial correlation vine approach the parameterization of the realized covariance matrices may...
edge weights based on empirical means

edge weights based on EWMA with $\lambda = 0.995$

edge weights based on EWMA with $\lambda = 0.98$

**Figure 8:** Illustration of the dynamic parameterization of the series of realized correlation matrices in terms of partial correlation vines. The horizontal time axis states the prediction month for which the associated R-vine structure was chosen. For the latter the selected pairs in $T_1$ are illustrated as black squares. The green squares indicate selectable pairs allowed by the proximity condition (which does not trigger in $T_1$). In the first row the average correlations used for R-vine structure selection are the empirical mean of the training set data, in the second, third row respectively, they are exponentially weighted moving averages with $\lambda = 0.995$, $\lambda = 0.98$ respectively.

dynamically change over time. Based on the current training set the R-vine structure selection method as described in Section 3.2 is applied to choose the subset $C_{W_1}(V_{6,M})$. Depending on how the average correlation matrix used for R-vine structure selection is calculated the parameterization is more or less sensitive to market developments. In Figure 8, the selected first trees of each of the 75 time windows are shown. Each time window is labeled by its prediction month. Black squares indicate pairs chosen by the R-vine structure selection method. In the first row empirical means of the pairwise standard correlations are considered. In the second row, resp. third row, exponentially weighted moving averages based on $\lambda = 0.995$, resp. $\lambda = 0.98$, are shown. While in case of empirical means all days of the two training years are of equal weight, for $\lambda = 0.995$, resp. $\lambda = 0.98$, the six most recent months, resp. one and a half months, already contribute half of the information for average calculation.

As to be expected the parameterization which is most sensitive to market changes varies the most over time. Changes in the selected pairs, e.g. for the prediction months in mid 2004 or at the beginning of the financial crisis, are observed earliest. Nevertheless, for all three scenarios the parameterizations are quite stable and may be divided into the periods from...
February 2002 – August 2006

February 2002 – August 2006, September 2006 – July 2007 and August 2007 – July 2008. For these periods Figure 9 illustrates the first tree $T_1$ of the predominantly chosen R-vine structures. Until August 2006 pairwise correlations including Citigroup (C) and General Electric (GE) seem to be most pronounced. While Citigroup plays a key role within the financial sector, General Electric 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 Citigroup becomes the core of the R-vine structure, i.e. the node 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 and the other market participants seem to tighten. This results in a predominantly chosen R-vine structure, where except for General Electric all pairwise correlations with JP Morgan are modeled. Note that in 2007 JP Morgan replaced Citigroup as the biggest US-bank in terms of revenues.

Clearly, the parameterization based on partial correlation vines gives interesting insights into market activities over time and therefore has a clear practical interpretation. Further, the ability to adapt to market changes over time ensures the goal of parsimonious time series modeling, which is based on capturing strong (partial) correlations in lower tree levels.

### 4.3 Estimation results

For each time window application of the partial correlation vine approach first results in selection of 15 (partial) correlation time-series and 6 variance time-series (see Section 3.2), i.e. in a 21-dimensional time-series estimation problem. In a copula based approach appropriate univariate time-series models for marginal modeling need to be found first (see Section 3.3). 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 modeling and forecasting purposes. To support this presumption for each period within the moving window approach time-series models of different complexity were fitted to the log-transformed variance time-series of JPM and for exemplary Fisher-z transformed (partial) correlation time-series. Also, the Cholesky elements within the Cholesky decomposition based benchmark model were investigated. Besides simply considering the mean value over time, as time-series models an AR(1) model, a basic HAR model as well as a HAR model including a GARCH(1,1) component with normal residuals (HARN) and with SGED residuals (HARSGED) were investigated. Figure 10, resp. Figure 11, illustrates on its left time-series of so-obtained BIC values, resp. values of the root mean squared errors.
Figure 10: BIC values for each of the 75 time windows within the moving window approach assuming univariate time-series models of increasing complexity (mean, AR(1), HAR, HARN, HARSGED). In the first row, estimation results for the log-transformed time-series of JPM are shown. In rows 2 to 4 exemplary Fisher-z transformed (partial) correlation time-series of increasing order are considered. An exemplary Cholesky element is given in the last row.
Figure 11: Root mean squared errors (RMSE) of forecasts made for each of the 75 time windows within the moving window approach assuming univariate time-series models of increasing complexity for predicting (mean, AR(1), HAR, HARN, HARSGED). In the first row, estimation results for the log-transformed time-series of JPM are shown. In rows 2 to 4 exemplary Fisher-z transformed (partial) correlation time-series of increasing order are considered. An exemplary Cholesky element is given in the last row.
(RMSE) for forecasts based on each of the univariate time-series model fits, and the corresponding boxplots on its right. Figure 1 and Figure 2 in the supplementary material show results for AIC values and mean absolute errors (MAE). While for the variance time-series and the standard correlation time-series in $T_1$ the HAR model and its extended versions clearly perform best, simply using the mean partial correlation value as fit and forecast is already sufficient for the example time-series in $T_4$. This holds not true for the illustrated partial correlation time-series associated with $T_3$. Here, the dispersion of both the BIC and the RMSE in case of using the mean or the AR(1) is quite high. Recall that given the R-vine structure selection method (Section 3.2) this time-series would not be selected for modeling purposes. Further, for the studied Cholesky element the HARGSED model, i.e. the most complex model, performs best. This result was already stated in Brechmann et al. (2015). In particular, it holds true for all 21 Cholesky elements, i.e. within the parameterization based on the Cholesky decomposition the information content is equally distributed among all Cholesky elements resulting in a homogeneous high data complexity.

Given these findings within the partial correlation vine approach we use HARN models for modeling the realized variance time-series and the realized correlation time-series appearing in $T_1$. For the realized partial correlation time-series in $T_2$ and $T_3$ basic HAR models are assumed. In $T_4$ and $T_5$ simply the mean value over the underlying training set is taken. Referring to Brechmann et al. (2015) and based on our own investigation HARGSED models are needed for all 21 Cholesky element time-series.

| PCV1 | PCV2 | PCV3 | PCV4 | PCV5 | PCV6 | PCV7 | PCV8 | PCV9 | PCV10 | PCV11 | PCV12 | PCV13 | PCV14 | PCV15 | PCV16 | PCV17 | PCV18 | PCV19 | PCV20 | PCV21 |
|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| 0.38 | 0.33 | 0.30 | 0.27 | 0.088| 0.022| 0.061| 0.033| 0.06  | 0.041 | 0.069 | 0.041 | -0.017| 0.078 | 0.23 | 0.25 | 0.16 | 0.19 | 0.18 | ... | 0.31 |
| PCV17| 0.34 | 0.33 | 0.34 | 0.49 |
| PCV18| 0.34 | 0.29 | 0.36 |
| PCV19| 0.29 | 0.35 |
| PCV20| 0.33 |
| PCV21| 0.33 |

**Figure 12:** Exploratory data analysis for the pairwise dependencies of the 21-dimensional pseudo u-data estimated for the period July 2006 to July 2008. Pairwise contour plots with normalized margins, histograms and pairs plots with empirical Kendall’s $\tau$ values are shown. The first five variables PCV1 to PCV5 correspond to data associated with the standard correlations in $T_1$, variables PCV6 to PCV9 correspond to data associated with the partial correlations in $T_2$, etc. Variables PCV16 to PCV21 correspond to data associated with the six variances.
For the residuals obtained through application of the univariate time-series models interest is on the pairwise cross-sectional association. 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 marginal fits, pseudo copula data as described in Section 3.3 are calculated. Figure 12 illustrates histograms of the so-obtained marginal pseudo u-data on its diagonal, pairwise contour plots with standard normal margins in the lower left corner and pairs plots with corresponding Kendall’s $\tau$ values in the upper right corner. Only dependencies between u-data associated to the variances (last six variables) and the selected standard correlations (first 5 variables) are significant with Kendall’s $\tau$ values ranging from 0.2 to 0.5. Dependencies associated to partial correlations are neglectable. Based on these findings and in favor of model parsimony, we assume independence for all pairs.

Based on the above findings Table 3 summarizes the model specifications together with the number of parameters, which need to be estimated for both the partial correlation vine approach and the Cholesky decomposition based approach. Clearly, due to the resulting inhomogeneous data complexity the parameterization based on partial correlation vines is preferable in terms of model parsimony.

Table 3: Summary of model specifications for the partial correlation vine approach (PCV) and the Cholesky decomposition based estimation model with the total number of parameters to be estimated.

| approach | variable type | #time-series | prediction model | #parameters |
|----------|---------------|--------------|-----------------|-------------|
| PCV      | variance      | 6            | HARN            | 42          |
|          | standard cors | 5            | HARN            | 35          |
|          | 1st order pcors | 4           | HAR             | 24          |
|          | 2nd order pcors | 3           | HAR             | 18          |
|          | 3rd order pcors | 2           | mean            | 1           |
|          | 4th order pcors | 1           | mean            | 1           |
|          | underlying copula |          | independence    | 0 / $\Sigma = 121$ |
| Cholesky | Cholesky elements | 21           | HARSGED         | 189         |
|          | underlying copula |            | independence    | 0 / $\Sigma = 189$ |

4.4 Forecasting performance

Based on the specified models out-of-sample one-day-ahead forecasts for the variances and (partial) correlations, resp. the Cholesky elements, are generated. Back-transformation of the predicted data via recursion formula (1) together with the predicted variances, resp. based on the Cholesky decomposition, gives forecasts for the realized covariance matrices.

To illustrate that the partial correlation vine approach is on target Figure 13 shows for the variance time-series of JPM (top panel), the covariance time-series for C and JPM (mid panel) as well as IBM and JPM (bottom panel) the full historical time-series from January 2002 until July 2008 together with the predicted values. Results for all other variances and covariance pairs look similar and are given in Figure 4 in the supplementary material. 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 both the Cholesky decomposition based and the partial correlation vine based approach. It is due to the high volatility of the true realized variances and covariances, while the predicted time-series rely on averaged historical data. Therefore, predictions are more stable (see Section 3.3). To further investigate the forecasting performance of the two approaches with regard to the single matrix elements, for each prediction month the RMSE is calculated. For the same examples as before, Figure 14 shows the time-series of the log-ratio
log(RMSE_{Cholesky} \ \text{\textbackslash} \ \text{RMSE}_{PCV}) on its left and corresponding boxplots on its right. Figure 5 in the supplementary material illustrates similar results for the remaining variance and covariance time-series. Note that by definition positive values of the considered log-ratio are in favor of the partial correlation vine approach. In economically quiet years, the slight fluctuations around zero show quite comparable forecasting performance of both modeling approaches. In times of financial turmoil, for almost all single elements there are clear positive peaks, which indicate better adaption to market changes in case of partial correlation vine based modeling. Here, the dynamic parameterization allows for more robust forecasts resulting in smaller and less divergent losses. Given that the Cholesky decomposition is unique for a fixed ordering of the variables within the underlying realized covariance matrices, marginal modeling of the Cholesky elements might be particularly challenging for certain time-periods. For each time-series, the median of the log-ratio is close to zero with a slight shift in favor of partial correlation vine based modeling. Several positive outliers confirm the good performance of the proposed modeling strategy. Finally, for each prediction day distances between the true realized covariance matrix and its partial correlation vine based forecast, resp. Cholesky decomposition based forecast, are compared considering the mean squared error for matrices, the Neumann divergence and the Bregman divergence. Figure 15 shows boxplots of the log-ratio log(distance_{Cholesky} \ \text{\textbackslash} \ \text{distance}_{PCV}) for the year 2005 on its left and for the period of July 2007 until July 2008 on its right. As for single elements the boxplots support the good performance of the partial correlation vine based approach with a positive median value and high positive outliers particularly in times of financial turmoil.

To conclude, the proposed forecasting approach yields good forecasts compared to the Cholesky decomposition based prediction model. In turbulent times, the advantages of dynamic parametrization become evident. Forecasts are more stable and at the same time reflect peaks in the original data. Recall that these results rely on simple and parsimonious time-series models for the majority of partial correlation time-series.

\textbf{Figure 13:} Daily realized variance time-series for JPM (1st row) and daily realized covariance time-series (2nd and 3rd ) together with the time-series of the corresponding daily forecasts.
Figure 14: Log-ratio \( \log(\text{RMSE}_{\text{Cholesky}}/\text{RMSE}_{\text{PCV}}) \) for forecasts made for each of the 75 time windows within the moving window approach. In the first row, forecasting results for the variance time-series of JPM are shown. In rows 2, resp. 3, covariance time-series of C and JPM, resp. IBM and JPM, are considered. On the left, time-series of the log-ratios are shown. On the right, corresponding boxplots are given.

Figure 15: Boxplots of the daily log-ratio \( \log(\text{distance}_{\text{Cholesky}}/\text{distance}_{\text{PCV}}) \). The mean squared error for matrices, the Neumann divergence and the Bregman divergence are considered as distance measures between the true realized covariance matrices and the partial correlation vine based matrix forecasts, resp. the Cholesky decomposition based matrix forecasts. Year 2005 and the period from July 2007 to July 2008 are considered separately.
5 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 valid parameterization tool. Along with a real data example we explore in detail the benefits of the proposed methodology as compared to a Cholesky decomposition based competitor model. Given the large number of R-vine structures for data transformation, we propose an R-vine structure selection method, which is completely based on historical information of the underlying data. It allows to dynamically change over time and therewith to adapt to market changes. The selection method is motivated by the practical interpretation of standard and partial correlations and allows for interesting insights in market behavior through the parameterization. High dependencies are captured by selected standard correlations and selected lower order partial correlations leaving higher order partial correlations, for which simple univariate time-series modeling is sufficient. As a consequence model parsimony is achieved. Studying the forecasting performance shows good and stable predictions compared to the benchmark model. The benefits of dynamic parametrization become most evident in times of financial turmoil.

In this work, focus is on the general introduction of the partial correlation vine approach and its strengths in terms of practical interpretation and model parsimony. In an ongoing project we study refinements of the proposed methodology by including flexible dependence models for the (partial) correlation time-series obtained after R-vine structure selection.

6 Acknowledgments

The authors gratefully acknowledge the support by Deutsche Forschungsgemeinschaft on the project ‘Vine copula based modeling and forecasting of multivariate realized volatility time-series’ [DFG CZ 86/4-1, OK 103/5]. Numerical calculations were performed on a Linux cluster supported by DFG grant INST 95/919-1 FUGG.

References

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 Labys, P. (2003). Modeling and forecasting realized volatility. Econometrica, 71(2):579–625.

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.

Barndorff-Nielsen, O. E. and Shephard, N. (2004). Econometric analysis of realized covariance: 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.

Brechmann, E. C., Heiden, M., and Okhrin, Y. (2015). A Multivariate Volatility Vine Copula Model. Econometric Reviews, page DOI:10.1080/07474938.2015.1096695.
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.

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.

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

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

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. 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.

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.

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

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, M. (1959). *Fonctions de répartition à n dimensions et leurs marges*. Université Paris 8.

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.

Yule, G. U. and Kendall, M. (1965). An introduction to the theory of statistics Charles Griffin & Co. *Belmont, CA*.

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.
Supplementary Material

A Skewed generalized error distribution (SGED)

The innovations $\epsilon_t$ of the GARCH (1, 1) component used for augmenting the basic HAR model can be assumed to follow a skewed generalized error distribution (SGED). We then consider

$$u_t = h_t \epsilon_t,$$
$$h_t^2 = \omega + \tilde{\alpha}_1 u_{t-1}^2 + \tilde{\beta}_1 h_{t-1}^2,$$
$$\epsilon_t \sim \text{SGED}(\mu, \sigma, \nu, \xi).$$

The skewed generalized error distribution is specified by the location parameter $\mu$, the scale parameter $\sigma$, the shape parameter $\nu$ and the skewness parameter $\xi$. Its density function is given by

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

with

$$C = \frac{\nu}{2\theta} \Gamma \left( \frac{1}{\nu} \right)^{-1},$$
$$\theta = \Gamma \left( \frac{1}{\nu} \right)^{1/2} \Gamma \left( \frac{3}{\nu} \right)^{-1/2},$$
$$\delta = 2\xi AS(\xi)^{-1},$$
$$S(\xi) = \sqrt{1 + 3\xi^2 - 4A^2\xi^2},$$
$$A = \Gamma \left( \frac{2}{\nu} \right) \Gamma \left( \frac{1}{\nu} \right)^{-1/2} \Gamma \left( \frac{3}{\nu} \right)^{-1/2}.$$

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

Figure 1: AIC values for each of the 75 time windows within the moving window approach assuming univariate time-series models of increasing complexity (mean, AR(1), HAR, HARN, HARSGED). In the first row, estimation results for the log-transformed time-series of JPM are shown. In rows 2 to 4 exemplary Fisher-z transformed (partial) correlation time-series of increasing order are considered. An exemplary Cholesky element is given in the last row.
Figure 2: Mean absolute errors (MAE) of forecasts made for each of the 75 time windows within the moving window approach assuming univariate time-series models of increasing complexity for predicting (mean, AR(1), HAR, HARN, HARSGED). In the first row, estimation results for the log-transformed time-series of JPM are shown. In rows 2 to 4 exemplary Fisher-z transformed (partial) correlation time-series of increasing order are considered. An exemplary Cholesky element is given in the last row.
Figure 3: Exploratory data analysis for the pairwise dependencies of the 21-dimensional pseudo u-data estimated for the period July 2006 to July 2008 using the Cholesky decomposition based estimation method. Pairwise contour plots with normalized margins, histograms and pairs plots with empirical Kendall’s τ values are shown. The variables Chol1 to Chol21 correspond to data associated with the 21 Cholesky elements.
Figure 4: (part 1/3) Daily realized variance time-series and daily realized covariance time-series together with the time-series of the corresponding daily forecasts.
Figure 4: (part 2/3) Daily realized variance time-series and daily realized covariance time-series together with the time-series of the corresponding daily forecasts.
Figure 4: (part 3/3) Daily realized variance time-series and daily realized covariance time-series together with the time-series of the corresponding daily forecasts.
Figure 5: (part 1/3) Log-ratio $\log(\text{RMSE}_{\text{Cholesky}}/\text{RMSE}_{\text{PCV}})$ for forecasts made for each of the 75 time windows within the moving window approach. On the left, time-series of the log-ratios are shown. On the right, corresponding boxplots are given.
Figure 5: (part 2/3) Log-ratio $\log(RMSE_{\text{Cholesky}}/RMSE_{\text{PCV}})$ for forecasts made for each of the 75 time windows within the moving window approach. On the left, time-series of the log-ratios are shown. On the right, corresponding boxplots are given.
Figure 5: (part 3/3) Log-ratio \( \log(\text{RMSE}_{\text{Cholesky}} / \text{RMSE}_{\text{PCV}}) \) for forecasts made for each of the 75 time windows within the moving window approach. On the left, time-series of the log-ratios are shown. On the right, corresponding boxplots are given.