

n° 2014-46

**Dynamic Asset Correlations
Based on Vines**

B.POIGNARD ¹
J-D FERMANIAN ²

Les documents de travail ne reflètent pas la position du CREST et n'engagent que leurs auteurs.
Working papers do not reflect the position of CREST but only the views of the authors.

¹ ENSAE, CREST and University Paris-Dauphine.

² ENSAE and CREST. E-mail: Jean-David.Fermanian@ensae.fr

Dynamic Asset Correlations Based on Vines

Benjamin Poignard* and Jean-David Fermanian†

February 5, 2015

Abstract

We develop a new method for generating dynamics of conditional correlation matrices between asset returns. These correlation matrices will be parameterized by a subset of their partial correlations, whose structure will be described by an undirected graph called “vine”. Since such partial correlation processes can be specified separately, our approach provides very flexible and potentially parsimonious multivariate processes. We introduce the so-called “vine-GARCH” class of processes and describe a quasi-maximum likelihood (QML) estimation procedure. Compared to other usual techniques, particularly for the Dynamic Conditional Correlation family, inference is simpler and can be led equation per equation. We compare our models with some DCC-type specifications through some simulated experiments and we evaluate their empirical performances by exploiting a database of daily stock returns.

Keywords: Dynamic Conditional Correlations, Multivariate GARCH, Partial Correlations, Quasi Maximum Likelihood Estimator, Regular vine.

*Ensaie-Crest & University Paris-Dauphine

†Ensaie-Crest

1 Introduction

Correlations between assets are key inputs in finance. For instance, building portfolio strategies most often requires the estimation of the correlations between asset returns, together with their volatilities. Obviously, multiple univariate settings are strongly insufficient to capture all dependencies among the levels, the volatilities and the correlations of asset returns in a portfolio. Therefore, a multivariate setting is necessary for modeling the cross-sectional and temporal dependencies between most asset returns. It allows for developing relevant management tools, especially when the interactions between financial markets become stronger. This concerns areas such as asset pricing, portfolio allocation, hedging derivatives or risk management measurement.

The usual modeling approach is to rely on the specification of the first two moments of vectors of returns conditional on their past (and current market information possibly), i.e. to specify their conditional mean and their conditional variance-covariance matrix. Once this is done, some assumed vectors of innovations close the model specification. Indeed, a vector of N asset returns is decomposed commonly as

$$r_t = \mu_t + \epsilon_t,$$

where μ_t denotes the vector of expected returns conditional on all information at time t , and ϵ_t has a (conditional) zero mean. While the financial literature about asset return predictability tends to focus on the modeling of μ_t (see Campbell and Shiller 1989, Campbell and Viceira 2002 e.g.), the modeling of time-varying covariances or correlations concentrates on the conditional variance H_t of ϵ_t , which is the “detrended” asset return.

If the temporal dependence of volatilities and/correlations is well-known among practitioners and researchers, the methods differ when modeling it. The multivariate GARCH (MGARCH) and the multivariate stochastic volatilities models are the two main approaches: see the surveys of Bauwens et al. (2006) and Asai et al. (2006) respectively. Such approaches allow for generating sequences of variance-covariance matrices, and then provide the correlations of asset returns as a by-product. In financial econometrics, the most commonly used specifications of H_t are provided surely by MGARCH models, because they aim at capturing the dynamics of the conditional variances and covariances straight without additional sources of randomness. However, the number of MGARCH parameters often increases dramatically with the number of underlying assets. Therefore, some simplified MGARCH models have searched for parsimony to allow for simple estimation and interpretation, but at the price of an over-simplification sometimes. Besides, MGARCH models have to take into account the positive definiteness of the model variance-covariance matrices, a technical constraint that may induce complexities.

Inside the MGARCH family, let us cite the Baba-Engle-Kraft-Kroner (1995, denoted by

BEKK) model, which specifies the dynamics of the underlying variance-covariance matrices directly as a deterministic quadratic function of past returns. The number of parameters in this class of MGARCH is significant, $O(N^2)$ to be specific. Hence Engle et al. (1990) proposed a new parameterization of H_t , called Factor-GARCH model, following the intuition that comovements of the asset returns are driven by a small number of common underlying variables. As a by-product, conditional correlations may be obtained in any MGARCH models, but their expressions are not intuitive or easily explicable. Indeed, they are complex nonlinear functions of observations and coefficients.

Actually, other specifications may be considered when focusing on conditional correlations. Intuitively, univariate GARCH dynamics may be chosen to get conditional variance processes. Then, based on these dynamics, a correlation process (R_t) could be built. This was the way proposed by Engle (2002) with the Dynamic Conditional Correlation (DCC). He extended the Constant Conditional Correlation (CCC) of Bollerslev (1990) by questioning the assumption that conditional correlations are constant. But to cope with the positive definiteness of R_t , for all t , DCC-type models have to rely on a normalization of the matrices they generate. This has been a source of difficulties and criticism (see Caporin and McAleer 2013), in particular to obtain a sound theory for inference. Note that this theoretical gap has been filled recently by Fermanian and Malongo (2013), who have exhibited some conditions for stationarity or ergodicity of DCC models. Moreover, although these families may allow for generating high-dimensional correlation matrices, estimation and forecasting are clearly challenging in their full form, for the same reason as MGARCH. Consequently, several attempts have tried to reduce strongly the number of parameters, notably the scalar DCC processes of Engle and Sheppard (2001), the Flexible DCC model of Billio and Caporin (2006), among others. But the ability of the latter models to capture complex and rich dynamics of heterogeneous series is limited, particularly for diversified portfolios.

Potentially, MGARCH models in general and DCC-type models in particular should be able to disentangle the volatility-correlation puzzle and to answer the following questions: does the volatility of an asset return lead the volatility of another one? Is the volatility of an asset transmitted to another asset through its conditional variance or its conditional covariance and/or correlation? Do correlations tend to get stronger during crisis? How do they impact portfolio strategies? Etc. Several papers try to answer such questions: among a lot of papers, see Bauwens and Otranto (2013) or Fermanian and Malongo (2014) recently. Nonetheless, the discussions around correlations often remain fragile and partly “black-box”, because neither MGARCH or DCC-type models work with correlations directly by as a subproduct. Indeed, the former ones set covariances when the latter ones depend on a normalization stage.

In this paper, we propose to circumvent the problem with another approach using partial correlations. This approach tends to be both parsimonious and flexible, and will specify some correlation and partial correlation dynamics directly. Any $N \times N$ correlation matrix

may be described by $N(N - 1)/2$ partial correlations. Lewandowski, Kurowicka and Joe (2009) explained how to deduce a correlation matrix from a partial correlation matrix (or the opposite), through an iterative algorithm. With such techniques, once the indices of a family of partial correlations is chosen conveniently, a “true” correlation matrix is generated whatever the values of these partial correlations. This property will be crucial here: by producing univariate dynamics of partial correlations independently, we obtain sequences of correlation matrices without the need for any normalization stage, contrary to DCC-type models.

An important practical question will be to choose the indices of the relevant partial correlations. Kurowicka and Cooke (2006) showed that the partial correlations of a random vector can be viewed as the output of vine trees. These objects are sets of connected undirected trees. They have been discovered recently due to their ability of building high-dimensional distributions thanks to a set of bivariate copulas (one copula per node of the graph) and conditional marginal cdfs’. See Aas et al. (2006) for an introduction. Here, we develop a class of MGARCH models based on regular vines, the so-called “vine-GARCH” models. The latter models are flexible enough by allowing independent specifications/estimation of partial correlation processes. It is also parsimonious as one can set constraints at any level of the vine tree without altering other correlations.

The rest of this paper is organized as follows: Section 2 develops some basic definitions/properties of trees, vines, partial correlations and the way they will be relevant for constructing nonnegative definite matrices. The vine-GARCH and its competing models are presented in Section 3. In Section 4, we detail the statistical inference of our new models by a QML procedure. Section 5 contains an empirical study with simulated data and a database of stock returns, and then we conclude the study.

The next section emphasizes how to specify a relevant set of partial correlations by considering a graphical approach based on vines.

2 Vines and partial correlations

First, we introduce some definitions from graph theory that will be used later in this paper.

2.1 Preliminary definitions

Graph theory was developed for resolving games. For instance, the Hamilton game consists of a map where you have points, representing a city, joined by edges, representing a path. The traveler should pass through all the cities but only once. Following the definitions of Napoles-Morales (2009, 2010), this game can be viewed as an *undirected graph*, denoted by $G = (N, E)$, consisting in a non empty set N of nodes and a possibly empty set E of edges. Each element of E is an *unordered* pair (α_1, α_2) , where α_1 and $\alpha_2 \neq \alpha_1$ belong to N . If the pair (α_1, α_2)

is *ordered*, the graph G is a *directed* graph, and, for (α_1, α_2) , this means there is a relation $\alpha_1 \rightarrow \alpha_2$. Then, α_1 is called a *parent* node and α_2 is a *child* node. If $N = \{1, 2, \dots, n\}$, we speak of a *labeled* graph.

Definition. The cardinality of N is called the *order* of the graph. If the pair (α_1, α_2) belongs to E then the two nodes α_1 and α_2 are *adjacent* and each one is incident with the pair (α_1, α_2) . The *degree* of a node is the number of edges incident with it.

Definition. A *path* of length n from α to β is a sequence $\alpha = \alpha_0, \dots, \alpha_n = \beta$ of distinct nodes such that $(\alpha_{i-1}, \alpha_i) \in E$ for $i = 1, \dots, n$. A *cycle* is a path such that $\alpha = \beta$. If every pair (α_{i-1}, α_i) in a cycle of a directed graph is ordered as in E then it is a *directed* cycle. Otherwise, it is an *undirected* cycle.

A *tree* is a particular case of a graph: it is an undirected and acyclic graph. It does not impose any path (i.e. undirected) between the nodes nor cycle (starting from point a and coming back). The formal definition follows:

Definition. $T = (N, E)$ is a tree with nodes N and edges E if E is a subset of unordered pairs of N with no cycle and there is a path between each pair of nodes. That is, there does not exist a sequence a_1, \dots, a_k for $k > 2$ of elements of N such that

$$(a_1, a_2) \in E, \dots, (a_{k-1}, a_k) \in E, (a_k, a_1) \in E.$$

And for any $a, b \in N$, there exists a sequence c_2, \dots, c_{k-1} of elements of N such that

$$(a, c_2) \in E, (c_2, c_3) \in E, \dots, (c_{k-1}, b) \in E.$$

T is called a *connected tree* if there is a path between all nodes of T .

Second, we use the latter concepts to detail vines, some particular graphical structures that are key in this paper.

2.2 Vines

Vines are undirected graphs that allow nice specifications of multivariate joint distributions. To be specific, a vine on n variables is a nested set of connected trees T_1, \dots, T_{n-1} where the edges of tree T_j are the nodes of tree T_{j+1} , $j = 1, \dots, n - 2$. A *regular vine* (R-vine) on n variables is a vine in which two edges in tree T_j are joined by an edge in tree T_{j+1} only if these edges share a common node, for any $j = 1, \dots, n - 2$.

Definition. $V(n)$ is a labeled *regular vine* on n elements if:

1. $V(n) = (T_1, T_2, \dots, T_{n-1})$.

2. T_1 is a connected tree with nodes $N_1 = 1, 2, \dots, n$ and edges E_1 . For $i = 2, \dots, n - 1$, T_i is a connected tree with nodes $N_i = E_{i-1}$, and the cardinality of N_i is $n - i + 1$.
3. If a and b are nodes of T_i connected by an edge in T_i , where $a = \{a_1, a_2\}$ and $b = \{b_1, b_2\}$, then exactly one of the a_i equals one of the b_i . This is the *proximity* condition.

We will consider only regular vines in this paper, and the properties we state hereafter are true for such vines, implicitly. A regular vine is called a *canonical vine* (C-vine) if each tree T_i has a unique node of degree $n - i$, i.e. a node with maximum degree. A regular vine is called a *D-vine* if all nodes in T_1 have degree not higher than 2. There are $n(n - 1)/2$ edges in a regular vine on n variables. An edge in tree T_j is an unordered pair of nodes of T_j , or equivalently, an unordered pair of edges of T_{j-1} .

In the case of vines, it makes sense to specify the links between the nodes of different T_j .

Definition. If node e is an element of node f , e is an *m-child* of f and we denote $e \in f$. If e is reachable from f via the *membership relation* $e \in e_1 \in e_2 \in \dots \in f$, e is an *m-descendent* of f .

A regular vine is a way of identifying a set of conditional bivariate constraints. The conditional bivariate constraints associated with each edge are determined as follows: the variables reachable from a given edge via the membership relation are called the *constraint set* of that edge. When two edges are joined by an edge of the next tree, the intersection of the respective constraint sets are the *conditioning variables*, and the symmetric differences of the constraint sets are the *conditioned variables*. With the notations of point 3 of the previous definition, at tree T_i , say $a_1 = b_1$, and a_1 is a common element of a and b . This means that, at tree T_{i+1} , a_1 enters the conditioning set of (a_2, b_2) (see next definition). Thus we define the conditioning and conditioned sets formally as follows:

Definition. For $e \in E_i$, $i \leq n - 1$, the *constraint set* associated with e is the complete union of the elements in $\{1, \dots, n\}$ that are reachable from e by the membership relation. It is denoted by U_e^* .

Definition. For $i = 1, \dots, n - 1$, if $e \in E_i$, it connects two elements j and k in N_i and it can be written $e = \{j, k\}$. The *conditioning set* associated with e is

$$L_e := U_j^* \cap U_k^*,$$

and the *conditioned set* associated with e is a pair

$$\{C_{e,j}, C_{e,k}\} := \{U_j^* \setminus L_e, U_k^* \setminus L_e\}.$$

Obviously, since the edges of a given tree T_j are the nodes of T_{j+1} , the same concepts of constraint/conditioning/conditioned sets apply to the nodes in a vine. It can be proved that,

in a regular vine, two different edges have different constraint sets. Moreover, if $e \in E_i$, then $\#U_e^* = i + 1$, $\#L_e = i - 1$ and there are necessarily two elements in its conditioned set: see Bedford and Cooke (2002).

In a regular vine, the edges of T_{m+1} (equivalently the nodes of T_{m+2}) will be denoted by $e = (a_j, a_k | b_1, \dots, b_m)$, where a_j, a_k and the $b_l, l = 1, \dots, m$ are different elements in $\{1, \dots, n\}$. This notation means that the conditioning set of e is $L_e = \{b_1, \dots, b_m\}$, and the conditioned set of e is $\{a_j, a_k\}$. Note that the conditioning sets are the same for all the nodes of a tree in a C-Vine. Moreover, the conditioning sets of the nodes in a D-vine are all different.

To give an example, consider 5 variables (X_1, \dots, X_5) . We consider here a C-Vine, which is built as follows:

- 1 is the root of the tree T_1 , i.e. its edges are $(1, 2)$, $(1, 3)$, $(1, 4)$ and $(1, 5)$. With our previous notations, we have $L_1 = L_2 = L_3 = L_4 = L_5 = \emptyset$. Moreover, $\{1\}$ constitutes the conditioning sets of E_1 (or N_2 , equivalently).
- In the tree T_2 , we choose (arbitrarily) 2 as the common element of all the edges, that become $(2, 3|1)$, $(2, 4|1)$ and $(2, 5|1)$. By the proximity condition, 1 and 2 enter the conditioning sets at tree T_3 .
- The same logic holds for element 3 in the tree T_3 . Its nodes are the edges of T_2 and its edges are $(3, 4|1, 2)$ and $(3, 5|1, 2)$.
- Finally, there remain two nodes for T_4 . Its single edge is $(4, 5|1, 2, 3)$.

The latter C-vine can be visualized on Figure 1 in the Appendix. For the sake of illustration, an example of a five order D-vine (resp. R-vine) is shown on Figure 2 (resp. Figure 3) too.

Let us finish this rather theoretical subsection with some useful results about R-vines.

Lemma 2.1. (Bedford, Cooke (2002))

Let a regular vine on n variables. Then,

1. the total number of edges is $n(n - 1)/2$,
2. each conditioned set is a doubleton, each pair of variables occurs exactly once as a conditioned set,
3. if two edges have the same conditioning set, then they are the same edge.

Lemma 2.2. (Kurowicka, Cooke (2006))

For any node \mathcal{N} in a regular vine, if the variable i is a member of the conditioned set of \mathcal{N} , then i is a member of the conditioned set of exactly one of the m -children of \mathcal{N} , and the conditioning set of an m -child of \mathcal{N} is a subset of the conditioning set of \mathcal{N} .

Lemma 2.3. (*Napoles-Morales (2010)*)

If the first tree of a vine on n nodes has one node with maximal degree, then the number of labeled regular vines possible with this tree equals the number of labeled regular vines on $n - 1$ nodes.

Note this is the case for C-vines.

Lemma 2.4. (*Napoles-Morales (2010)*)

If the first tree of a vine on n nodes has $(n - 2)$ nodes with degree 2, then the number of regular vines possible with this tree equals 1.

Note this is the case for D-vines.

Lemma 2.5. (*Napoles-Morales (2010)*)

The number of C-vines on n nodes equals the number of D-vines on n nodes and is $n!/2$.

Proof. For C-vines, observe that there are n possible labeled trees on n nodes for which a single node has maximal degree. Once the first tree has been fixed any of the $(n - 1)$ edges may be chosen so as to construct any of the $(n - 1)$ possible labeled trees on $(n - 1)$ nodes for which a single node has maximal degree. Any of these would preserve regularity. The same argument holds for all other trees on the vine until two edges need to be connected as nodes in T_{n-1} . Hence there are $n(n - 1)(n - 2)(n - 3) \cdots 3 = n!/2$ C-vines on n nodes. For D-vines observe that from lemma 2.4, $T_1 \in V$ completely determines the vine. And since there are $n!/2$ ways of choosing it, the result follows. \square

Lemma 2.6. (*Napoles-Morales (2010)*)

The total number of regular vine is $C_n^2 \times (n - 2)! \times 2^{C_n^2}$ for n nodes.

Now, let us show how we can link such vines with some subsets of the partial correlations that are associated to a random vector.

2.3 Partial correlations

Let X_i be a n -dimensional random variable, $n \geq 2$, with zero mean. Let $b_{1,n|2,\dots,n-1}, \dots, b_{n-1,n|1,\dots,n-2}$ be the solutions of the minimization program

$$\arg \min_{b_1, \dots, b_{n-1} \in \mathbb{R}} E \left[(X_n - b_1 X_1 - b_2 X_2 - \dots - b_{n-1} X_{n-1})^2 \right].$$

Definition. The *partial correlation* between X_n and X_{n-1} given X_1, \dots, X_{n-2} is

$$\rho_{n,n-1|1,\dots,n-2} = \text{sgn} (b_{n,n-1|1,\dots,n-2}) (b_{n,n-1|1,\dots,n-2} b_{n-1,n|1,\dots,n-2})^{1/2}.$$

By changing the indices, we define similarly the partial correlation between X_i and X_j knowing X_k , $k \in \{1, \dots, n\} \setminus \{i, j\}$, or even knowing any subset of variables X_k where k belongs to $L \subset \{1, \dots, n\}$. In the latter case, the corresponding partial correlation is denoted by $\rho_{i,j|L}$.

More intuitively, a partial correlation can be defined as partial regression coefficients. For instance, $\rho_{n,n-1|1,\dots,n-2}$ corresponds to the correlation between the orthogonal projections of X_n and X_{n-1} on the orthogonal of the space spanned by (X_1, \dots, X_{n-2}) . Equivalently, the partial correlation can be defined as

$$\rho_{n,n-1|1,\dots,n-2} = -\frac{K_{n,n-1}}{\sqrt{K_{n,n}K_{n-1,n-1}}},$$

where $K_{i,j}$ denotes the (i, j) cofactor of the correlation matrix of (X_1, \dots, X_n) .

More generally, $\rho_{i,j|L}$ corresponds to the correlation between the orthogonal projections of X_i and X_j on $\langle X_k, k \in L \rangle^\perp$, the orthogonal of the subspace generated by $\{X_k, k \in L\}$. When L is empty, we recover usual correlations, i.e. $\rho_{i,j|\emptyset} = \rho(X_i, X_j) := \rho_{i,j}$, the usual correlation between X_i and X_j .

Note that, if the random vector (X_1, \dots, X_n) is normal and is specified by its full-rank variance covariance matrix, then its partial correlations correspond to some conditional correlations.

Interestingly, partial correlations can be computed from (usual) correlations with a recursive formula. let X_1, \dots, X_n be random variables, (i, j, k) be a set of distinct indices, and L be a set of indices (possibly empty) that is disjoint from (i, j, k) . Following Lewandowski, Kurowicka and Joe (2009), the partial correlation of X_i and X_j given $\{X_k; k \in L\}$ satisfies

$$\rho_{i,j|k,L} = \frac{\rho_{i,j|L} - \rho_{i,k|L}\rho_{j,k|L}}{\sqrt{(1 - \rho_{i,k|L}^2)(1 - \rho_{j,k|L}^2)}}. \quad (1)$$

Assume we know the usual correlations $\rho_{i,j}$, for any couple (i, j) , $i \neq j$. Any partial correlation can be calculated by invoking (1) several times with increasing subsets L . Actually, the opposite is true if we start from a convenient subset of partial correlations, as we explain now.

The edges of a regular vine may be associated with the partial correlations of a n -dimensional random vector in the following way: for $i = 1, \dots, n-1$, consider any $e \in E_i$, the set of edges at tree T_i . Let $\{j, k\}$ be the two conditioned variables of e , and L_e its conditioning set. We associate the partial correlation $\rho_{j,k|L_e}$ to this node. Kurowicka and Cooke (2006) call this structure a *partial correlation vine*, that is simply a R-vine for which any node is associated to a number in $] -1, 1[$. Actually, all correlation vines are generated by setting a R-vine on n variables, and by assigning a partial correlation, say $\rho_{\cdot,\cdot|e}$ with any $e \in \bigcup_{i=1}^{n-1} E_i$, arbitrarily taken in $] -1, 1[$. In particular, partial correlation vines allow for a nice parametrization of spherical distributions.

Theorem 2.7. (*Bedford, Cooke (2002)*)

For any regular vine on n elements, there is a one-to-one mapping between the set of $n \times n$ positive definite correlation matrices and the set of partial correlation specifications of the vine.

And the formulas (1) above allow the building of a true $n \times n$ correlation matrix based on any set of $n(n-1)/2$ partial correlations that are deduced from a regular vine (see Kurowicka and Cooke, 2003).

To illustrate the latter property, let us detail the type of calculations we have to lead to get usual correlations from a subset of partial correlations. Assume we know the partial correlations that correspond to any node of the previous C-vine. In other words, we know $\rho_{12}, \rho_{1,3}, \rho_{1,4}, \rho_{1,5}, \rho_{2,3|1}, \rho_{2,4|1}, \rho_{2,5|1}, \rho_{3,4|12}, \rho_{3,5|1,2}$ and $\rho_{4,5|1,2,3}$. We can deduce from these quantities all the usual correlations by applying equation (1) several times. For instance, we would like to compute the usual correlation between 4 and 5. The idea is to start from the node in the vine where $\{4, 5\}$ is the conditioned set. This occurs only once, due to Lemma 2.1. We observe that the node $(4, 5|1, 2, 3)$ appears in T_4 , and we “go down” the vine:

$$\rho_{4,5|1,2,3} = \frac{\rho_{4,5|1,2} - \rho_{4,3|1,2}\rho_{5,3|1,2}}{\sqrt{(1 - \rho_{4,3|1,2}^2)(1 - \rho_{5,3|1,2}^2)}}.$$

The index 3 was taken out because the conditioning set at tree T_3 is $(1, 2)$. Then, we have

$$\rho_{4,5|1,2} = \rho_{4,5|1,2,3}\sqrt{(1 - \rho_{4,3|1,2}^2)(1 - \rho_{5,3|1,2}^2)} + \rho_{4,3|1,2}\rho_{5,3|1,2}$$

This quantity can be calculated because $\rho_{4,3|1,2}$ and $\rho_{5,3|1,2}$ are given in tree T_3 . But $\rho_{4,5|1,2}$ must also satisfy:

$$\rho_{4,5|1,2} = \frac{\rho_{4,5|1} - \rho_{4,2|1}\rho_{5,2|1}}{\sqrt{(1 - \rho_{4,2|1}^2)(1 - \rho_{5,2|1}^2)}}.$$

Again we can take 2 out by calculating

$$\rho_{4,5|1} = \rho_{4,5|1,2}\sqrt{(1 - \rho_{4,2|1}^2)(1 - \rho_{5,2|1}^2)} + \rho_{4,2|1}\rho_{5,2|1},$$

and we invoke finally

$$\rho_{4,5|1} = \frac{\rho_{4,5} - \rho_{4,1}\rho_{5,1}}{\sqrt{(1 - \rho_{4,1}^2)(1 - \rho_{5,1}^2)}}.$$

By calling the previous relations, we get the usual correlation between 4 and 5.

$$\rho_{4,5} = \rho_{4,5|1}\sqrt{(1 - \rho_{4,1}^2)(1 - \rho_{5,1}^2)} + \rho_{4,1}\rho_{5,1}$$

In the case of a R-vine, such recursive computations yield an algorithm to deduce the usual correlation matrix from the set of the partial correlations associated to a partial correlation vine.

Definition. Let a vine $V(n) = (T_1, T_2, \dots, T_{n-1})$. The set of (usual and) partial correlations associated to this vine is denoted by $\tilde{C}_{V(n)} := (C(T_1), C(T_2), \dots, C(T_{n-1}))$.

Since $\tilde{C}_{V(n)}$ enables to compute all the set of usual correlations $R := \{\rho_{i,j}; i, j = 1, \dots, n; i < j\}$, we denote by $R(\tilde{C}_{V(n)})$ the values the correlations that are deduced from $\tilde{C}_{V(n)}$.

We now turn to the significant results that ensure the positive definiteness of the correlation matrices when using vine representations. By recalling Equation (1), the following result makes sure that any correlation computed from arbitrary partial correlations (belonging to $] - 1, 1[$, obviously) is still an element in $] - 1, 1[$.

Lemma 2.8. (*Kurowicka, Cooke (2006)*)

If $z, x, y \in] - 1, 1[$, then also $w \in] - 1, 1[$ with

$$w = z\sqrt{(1-x^2)(1-y^2)} + xy.$$

The next result ensures the positive definiteness of the matrices $[\rho_{i,j}]$ when deduced from any set of partial correlations. This is a key result allowing the easy generation of sequences of correlation matrices. It will constitute an attractive feature of the vine-GARCH models we will introduce in Section 4.

Theorem 2.9. (*Kurowicka, Cooke (2006)*)

Let $D_n > 0$ be the determinant of the n -dimensional correlation matrix $\Sigma_n := [\rho_{i,j}]_{i,j=1,\dots,n}$. For any set of partial correlations generated by a regular vine,

$$D_n = \prod_{i=1}^{n-1} \prod_{e \in E_i} (1 - \rho_{j,k|L_e}^2), \quad (2)$$

where (j, k) and L_e are respectively the conditioned set and the conditioning set of an edge e .

Corollary 2.10. *Whatever the values of set of partial correlations generated by a regular vine on $\{1, \dots, n\}$, the associated matrix $[\rho_{i,j}]$ is nonnegative definite.*

Proof. By Theorem 2, D_n is nonnegative whatever the values of the partial correlations in $\mathcal{P}_n := \{\rho_{j,k|L_e}\}$, that induce the correlations $\rho_{i,j}$, $i, j = 1, \dots, n$. But the same result applies for every matrix Σ_k , $k = 1, \dots, n-1$ too. Indeed, given \mathcal{P}_n , we are able to calculate all the $\rho_{i,j}$, $i, j = 1, \dots, n$ (that belong to $[-1, 1]$ by Lemma 2.8), and then any set of partial correlations associated to any new vine on $\{1, \dots, k\}$, $k < n$ by invoking (1). And Theorem 2 can be applied to Σ_k . But a symmetrical matrix for which all the main block diagonal submatrices have nonnegative determinants is nonnegative. \square

In other words, whatever the values of the partial correlations $\tilde{C}_{V(n)} = (C(T_1), \dots, C(T_{n-1}))$ associated to a regular vine $V(n)$, we get a true correlation matrix with the coefficients $R(C(T_1), \dots, C(T_{n-1}))$.

The next subsection is related to the robustness of correlation calculations when some levels of partial correlations are set to zero inside a vine. It can be skipped at the first reading.

2.4 Truncation of vines

For the sake of parsimony, it will be interesting to cancel (or to leave constant, at least) all the partial correlations associated to a vine, after some level p . When zero partial correlations are assumed after the latter level, we would like to know whether the corresponding (usual) correlations depend on the trees $T_p, T_{p+1}, \dots, T_{n-1}$ that could be built above.

Definition. We say that a vine is p -VF (VF for vine-Free) if:

$$R(C(T_1), C(T_2), \dots, C(T_{n-1})) = R(C(T_1), C(T_2), \dots, C(T_{p-1}), C(T'_p), \dots, C(T'_{n-1}))$$

for any alternative vine $V'(n) := (T_1, T_2, \dots, T_{p-1}, T'_p, \dots, T'_{n-1})$, and when the partial correlations associated to the edges of T'_k , $k \geq p$, are zero.

In the definition above, if the set of alternative vines we consider is restricted to C-vines (resp. D-vines), we say that the initial vine is p -VF for C-vines (resp. D-vines). If a vine is p -VF, once the partial correlations are zero above the level p , the correlations are independent on the way this vine has been built from this level.

Property 2.11. A C-vine on n elements is p -VF for C-vines, for any $p \geq 1$.

Proof. Let us consider an arbitrary C-vine for which the associated partial correlations are zero from step p on. At tree T_1 , there is one element, say a_1 that is common to all the correlations/edges:

$$(a_1, a_2), (a_1, a_3), \dots, (a_1, a_n) \in E_1,$$

that are the nodes of T_2 . Let us denote by L_j the conditioning set that is common to all the edges in tree T_j , $j = 1, \dots, n - 1$. Obviously, $L_1 = \emptyset$. In T_2 , we build edges by connecting (a_1, a_2) with all the other nodes. Then the edges of this tree are $(a_2, a_3|a_1), \dots, (a_2, a_n|a_1)$. In tree T_3 , the conditioning set (of the edges) is $L_3 = (a_1, a_2)$. We choose a_3 as the common element of all the conditioned subsets, i.e. the edges of T_3 are $(a_3, a_i|a_1, a_2)$, where $i \notin \{1, 2, 3\}$. And we go on. The edges of T_k are $(a_l, a_k|L_k)$, $a_l \notin \{a_1, \dots, a_k\}$ and $L_k = (L_{k-1}, a_{k-1})$. Note that $L_1 = \emptyset$.

Now the partial correlations associated to the first $p - 1$ trees of all these C-vines will be assumed fixed and known. For such a C-vine, we assume the elements of $C(T_p), \dots, C(T_{n-1})$

are zero. Let us evaluate the correlation $\rho_{i,j}$ for an arbitrarily chosen C-vine in this family. From Proposition 2.1, (i, j) is the conditioned set of a unique edge $e_{i,j} := (i, j|L_k)$, that belongs to the tree T_k in this C-vine, for some $k = 1, \dots, n - 1$.

Case 1: $k < p$.

In this case, due to the C-vine assumption, i or j , say j , is the “common element” in tree T_k . In other words, $j = a_k$. The nodes i and j in T_1 may be connected to $(i, j|L_k)$ by a path that belongs to the first part (T_1, \dots, T_k) of the C-vine. By applying several times (1), we can calculate $\rho_{i,j}$ by involving only the partial correlations associated to the edges of $(T_1, \dots, T_k) \subset (T_1, \dots, T_p)$, that are fixed. To be specific, we apply the relation

$$\rho_{i,j|L_l} = \frac{\rho_{i,j|L_{l-1}} - \rho_{i,a_{l-1}|L_{l-1}}\rho_{j,a_{l-1}|L_{l-1}}}{\sqrt{\left(1 - \rho_{i,a_{l-1}|L_{l-1}}^2\right)\left(1 - \rho_{j,a_{l-1}|L_{l-1}}^2\right)}}, \quad (3)$$

for $l = k, k - 1, \dots, 1$. It can be checked easily that $\rho_{i,j}$ is a function of ρ_{i,a_1} , ρ_{j,a_1} , $\rho_{i,a_2|a_1}$, $\rho_{j,a_2|a_1}, \dots, \rho_{i,a_{k-1}|L_{k-1}}$, $\rho_{j,a_{k-1}|L_{k-1}}$ and $\rho_{a_i,a_j|L_k}$, that are all known.

Case 2: $k \geq p$.

We can start with the same logic as above. We apply (3) from L_k and backwards, choosing the common element that is related to the current tree (the denoted a_{l-1} in (3)). We observe that

$$\rho_{i,j|L_k} = \rho_{i,j|L_{k-1}} = \dots = \rho_{i,j|L_p} = 0.$$

The first (a priori) non zero partial correlation of this type is

$$\rho_{i,j|L_{p-1}} = \rho_{i,a_{p-1}|L_{p-1}}\rho_{j,a_{p-1}|L_{p-1}}.$$

But this relation is independent of the starting position of the node $e_{i,j}$ in the C-vine structure we have considered. Then, we are reduced the problem to the calculation of $\rho_{i,j}$ knowing $\rho_{i,j|L_{p-1}}$. This can be solved exactly as in Case 1 above. \square

Property 2.12. *A D-vine on n elements is p -VF for D-vines, and for any $p \geq 1$.*

Proof. The D-Vine structure implies that $(C(T_2), \dots, C(T_{n-1})) = (C(T'_2), \dots, C(T'_{n-1}))$. Indeed, the first tree T_1 determines the rest of the D-Vine structure. \square

Actually, once the first tree of a vine is the one of a D-vine, then the structure is uniquely stated and it remains no degree of liberty.

We have stated above the p -VF property, for any $p \geq 1$, and staying inside the families of C-or D-vines. Does this property hold for the general case, i.e. for regular vines? We conjecture that any regular vine is p th-VF, for every $p \geq 1$. And we prove the latter result is true in two particular cases.

Let regular vines with n elements. Suppose $\tilde{C}_{V(n)} := (C(T_1), C(T_2), \dots, C(T_{n-1}))$ where $(C(T_1), C(T_2), \dots, C(T_{p-1}))$ follows C-Vine rules until tree T_p , and then $(C(T_{p+1}), \dots, C(T_{n-1}))$ is built as a regular vine. This is called the CR-vine family of order p .

Theorem 2.13. *The CR-vines of order p are p -VF.*

Proof. The proof is very similar to the one of Property 2.11. But contrary to the standard C-vine case, the conditioning subsets may be different for different edges in any tree T_l , $l \geq p$.

Let us calculate $\rho_{i,j}$. In a R-vine that is C-vine until T_p , there exists an edge $e = (i, j|L_k(i, j))$ where the conditioned subset is $\{i, j\}$. Note that the conditioning set of e depends on (i, j) . If $k < p$, the same arguments of Property 2.11 apply. Otherwise, we can invoke the relation

$$\rho_{i,j|L_k(i,j)} = \frac{\rho_{i,j|L_{k-1}(i,j)} - \rho_{i,e_k|L_{k-1}(i,j)}\rho_{j,e_k|L_{k-1}(i,j)}}{\sqrt{\left(1 - \rho_{i,e_k|L_{k-1}(i,j)}^2\right)\left(1 - \rho_{j,e_k|L_{k-1}(i,j)}^2\right)}}, \quad (4)$$

for some conditioning set $L_{k-1}(i, j)$ in T_{k-1} , and where the node e_k has been chosen s.t. $(i, e_k|L_{k-1}(i, j))$ is an edge of T_{k-1} . This is always possible due to the proximity condition. But, by assumption, $\rho_{i,e_k|L_{k-1}(i,j)} = 0$. Then, $\rho_{i,j|L_{k-1}(i,j)} = 0$, and we go on until we reach the tree T_p . In the latter case, we have

$$\rho_{i,j|L_{p-1}(i,j)} = \rho_{i,a_{p-1}|L_{p-1}(i,j)} \cdot \rho_{j,a_{p-1}|L_{p-1}(i,j)},$$

that is known because the latter r.h.s. may be read in the first C-vine layer. Note that $L_{p-1}(i, j) = L_{p-1}$ is the (unique) conditioning set of T_{p-1} . Moreover, note that $L_p(r, s) = (a_{p-1}, L_{p-1})$ for any conditioned set (r, s) in T_p . Once $\rho_{i,j|L_{p-1}}$ is known, we get $\rho_{i,j}$ applying the formula of Cooke, Joe et Kurowicka backwards as usual, i.e. extracting the successive common elements $a_{p-1}, a_{p-2}, \dots, a_1$. So the result. □

3 A reminder about MGARCH and DCC-type models

3.1 Basics

The most usual approach for modeling multivariate dynamics of asset returns is given by the family of multivariate GARCH (MGARCH), as introduced by Bollerslev and al. (1988). By focusing on second-order conditional moments of random vectors, they extended the multivariate ARCH model of Engle (1982). They have been applied extensively in finance, particularly

in the framework of the Capital Asset Pricing Model and mean-variance portfolio optimization. A natural question was to design such models that would be sufficiently flexible for high dimension settings while being parsimonious. When dealing with correlation dynamics, several proposals emerged, notably the Dynamic Conditional Correlation model of Engle (2002) and its variations. For the moment, we introduce this usual framework and some notations that will be useful hereafter.

To set the ideas, let us consider a N -dimensional vectorial stochastic process $(r_t)_{t=1,\dots,T}$. We denote by θ the vector of the model parameters and decompose the stochastic process $(r_t)_{t=1,\dots,T}$ as the sum of conditional expected returns and a residual:

$$r_t = \mu_t(\theta) + \epsilon_t, \quad (5)$$

$$\epsilon_t = H_t^{1/2}(\theta) \eta_t. \quad (6)$$

Here, $\mu_t(\theta) = \mathbb{E}[r_t | \mathcal{F}_{t-1}] := \mathbb{E}_{t-1}[r_t]$, where \mathcal{F}_t denotes the market information until (and including) time t . We suppose $H_t(\theta) = \text{Var}(r_t | \mathcal{F}_{t-1}) := \text{Var}_{t-1}(r_t) = \text{Var}_{t-1}(\epsilon_t)$ is a $N \times N$ positive definite matrix. The series $(\eta_t)_{t \geq 0}$ is supposed to be a strong white noise, i.e. an independent and identically distributed sequence of random variables s.t. $\mathbb{E}[\eta_t] = 0$ and $\text{Var}(\eta_t) = I_N$. The model will be semi-parametric. Its specification is complete when the law of η_t is defined and the functional form of both $\mu_t(\theta)$ and $H_t(\theta)$ are specified. In this paper, we focus on the latter point. For convenience, we will denote $\mu_t(\theta) = \mu_t$ and $H_t(\theta) = H_t$.

To remove the first moment, we suppose simply that the conditional expected returns are modeled as AR(1), i.e. there exist Φ_0 a $N \times 1$ matrix and Φ_1 a $N \times N$ diagonal matrix s.t. $\mu_t(\theta) = \Phi_0 + \Phi_1 r_{t-1}$. Since we are interested in ϵ_t in this paper, we estimate μ_t by OLS and subtract it from r_t . Now, these estimated residuals will be considered as our observations. This is common practice in the MGARCH literature.

The matrix H_t represents the unobserved time-dependent dynamics of the variance-covariance matrix of the return process. A brute-force inference of the whole dynamics seems unfeasible when working in high-dimension. To avoid this problem, researchers prefer to split the problem into two simpler ones. They first specify N conditional volatility dynamics, often of the GARCH-type. Second, they specify the correlation dynamics between the individual conveniently normalized returns. We will follow this common practice.

3.2 The usual treatment of conditional correlations

Conditional variances will be denoted by $h_{ii,t}$ and the conditional correlations by $\rho_{ij,t}$, for $i, j = 1, \dots, N$, $i < j$. The conditional covariance between the assets i and j at time t is then $h_{ij,t} = \rho_{ij,t} h_{ii,t}^{1/2} h_{jj,t}^{1/2}$. In matrix notation, we write

$$H_t = D_t^{1/2} R_t D_t^{1/2},$$

where $D_t = \text{diag}(h_{11,t}, \dots, h_{NN,t})$ is the diagonal matrix of the conditional variances, and $R_t = (\rho_{ij,t})$, $1 \leq i \leq j \leq N$ is the matrix of the conditional correlations of r_t . By construction, R_t is the conditional covariance matrix of the vector of the standardized and demeaned returns $\epsilon_t = (\epsilon_{1,t}, \dots, \epsilon_{N,t})$ with $\epsilon_{i,t} = (r_{i,t} - \mu_{i,t}) / \sqrt{h_{ii,t}}$, also called “degarched” returns. The specification of H_t is divided into two independent parts: the conditional variance process on one side, and the conditional correlation process on the other side. Each of them depends on a set of parameters

$$\theta = (\theta_{vol}, \theta_{cor})' \in \Theta,$$

where θ_{vol} is the set of parameters determining the univariate volatility processes and θ_{cor} determines the correlation process. Θ is a compact set and the true parameter value $\theta_0 \in \Theta$. Typically, most authors assume that the univariate conditional variance processes are functions determined by their lag processes and the i -th element of ϵ_t only:

$$\begin{aligned} h_{ii,t} &= h_i(\theta_{vol,i}; \epsilon_{i,t-1}) \\ &= \varsigma_i + \kappa_i \epsilon_{i,t-1}^2 + \tau_i h_{ii,t-1}. \end{aligned} \tag{7}$$

Under the latter assumption, The process $(h_{ii,t})$ follows a GARCH(1,1) where $\theta_{vol,i} = (\varsigma_i, \kappa_i, \tau_i)' \in \mathbb{R}_+^3$ for all $i = 1, \dots, N$. Thus, $3N$ parameters need to be evaluated to describe all the individual volatility processes. To insure stationarity, the κ and τ coefficients have to satisfy the constraint $\kappa_i + \tau_i < 1$. We do not suppose the existence of spill-over effects between different assets, that is the presence of terms such as $h_{jj,t-1}$ or $\epsilon_{j,t-1}$ inside the $(h_{ii,t})$ dynamics. This usual assumption simplifies significantly the estimation of θ_{vol} by allowing an equation-by-equation inference procedure.

Several specifications exist for the (R_t) process. They have to challenge the positive definiteness of the correlation matrix and should not depend on too many parameters. Bollerslev (1990) devised the Constant Conditional Correlation model. But numerous empirical results tend to reject the constant correlation hypothesis. During financial crisis, it has been observed a strengthening of such correlations, in particular. Therefore, there was a need of more flexible models to tackle this issue. In particular, the Time-Varying Correlation of Tse and Tsui (2002) or the Dynamic Conditional Correlation (DCC) of Engle and Sheppard (2001) were the first answers. In this study, we consider the DCC family model as our benchmark.

3.3 DCC-GARCH models

The DCC model specifies dynamics of the variance covariance matrix of the de-garched returns ϵ_t directly. In its full form, called “Full DCC”, the model belongs to the MARCH family of

Ding and Engle (2002) and is specified as follows:

$$\begin{aligned} H_t &= D_t^{1/2} R_t D_t^{1/2}, \\ Q_t &= (\nu' - A - B) \odot S + A \odot \epsilon_{t-1} \epsilon_{t-1}' + B \odot Q_{t-1}, \\ R_t &= Q_t^{*-1/2} Q_t Q_t^{*-1/2}, \end{aligned}$$

where $Q_t = [q_{ij,t}]$ and $Q_t^* = \text{diag}(q_{11,t}, q_{22,t}, \dots, q_{NN,t})$. S , A and B are $N \times N$ symmetric matrices of unknown parameters. \odot denotes the usual Hadamard product of two identically sized matrices, which is computed simply by element-by-element multiplication. Following Ding and Engle (2002), if $(\nu' - A - B) \odot S$, A and B are positive semi-definite, then the matrix Q_t is positive semi-definite. If one of the matrices is positive definite, then Q_t is positive definite. S is positive definite matrix too. It is often approximated by the unconditional variance-covariance matrix of ϵ_t . This simplifies the estimation problem because this avoids an optimization on A , B and S simultaneously. This procedure is called *correlation targeting* or *correlation tracking*. However, the significant downside of the Full DCC model is its intractability: even after the targeting, the (Q_t) process encompasses $N(N - 1)$ coefficients. In most empirical studies, the scalar DCC-GARCH is considered instead, where A and B are replaced by non negative scalars α and β , $\alpha + \beta < 1$.

Between the Full DCC and the scalar DCC (2 coefficients, after correlation targeting), it remains some place for intermediate solutions. For instance, Billio and Caporin (2006) considered another parametrization, the Quadratic Flexible DCC (QFDCC), which reduces the size of the problem while remaining flexible. In the general form of a QFDCC model, the correlation driving process (Q_t) is defined as follows:

$$\begin{aligned} H_t &= D_t^{1/2} R_t D_t^{1/2}, \\ Q_t &= C' S C + A' \epsilon_{t-1} \epsilon_{t-1}' A + B' Q_{t-1} B, \\ R_t &= Q_t^{*-1/2} Q_t Q_t^{*-1/2}, \end{aligned}$$

where S , A , B and C are symmetric matrices. This model allows for interdependence across groups of assets, which is particularly adapted to cases where a portfolio is composed of different asset classes, or with assets coming from different industries/geographical areas. The QFDCC model provides positive definite correlation matrices if the eigenvalues of $A + B$ are less than one in modulus. This model is parsimonious when the matrices A , B and C are diagonal: this yields to a model with $3N$ unknown parameters, after correlation targeting. In our study, only the diagonal QFDCC model will be considered.

In the literature, the DCC-GARCH models with variance targeting are implemented generally by considering the matrix S as the unconditional covariance matrix of the standardized residuals. However, in the case of a scalar DCC, Aielli (2013) has shown that this way of

working produces biased estimates in general. The only case where the equality $S = \mathbb{E}[\epsilon_t \epsilon_t']$ holds is the one of constant correlations, i.e. $\alpha = \beta = 0$. Indeed, taken the expectation operator of the Q_t process in the scalar DCC case, we get

$$\mathbb{E}[Q_t] = (1 - \alpha - \beta)S + \alpha \mathbb{E}[\epsilon_t \epsilon_t'] + \beta \mathbb{E}[Q_t].$$

And $S = \mathbb{E}[\epsilon_t \epsilon_t']$ if and only if $\mathbb{E}[\epsilon_t \epsilon_t'] = \mathbb{E}[Q_t]$. But the following relationship holds:

$$\mathbb{E}[\epsilon_t \epsilon_t'] = \mathbb{E}[\mathbb{E}_{t-1}[\epsilon_t \epsilon_t']] = \mathbb{E}[R_t] = \mathbb{E}\left[Q_t^{*-1/2} Q_t Q_t^{*-1/2}\right].$$

Except in the constant correlation case, $Q_t \neq Q_t^{*1/2} Q_t Q_t^{*1/2}$. That is why Aielli reformulates the scalar DCC model as a *corrected* DCC (cDCC) as follows:

$$\begin{aligned} H_t &= D_t^{1/2} R_t D_t^{1/2}, \\ Q_t &= (1 - \alpha - \beta)S + \alpha \left\{ Q_{t-1}^{*1/2} \epsilon_{t-1} \epsilon_{t-1}' Q_{t-1}^{*1/2} \right\} + \beta Q_{t-1}, \\ R_t &= Q_t^{*-1/2} Q_t Q_t^{*-1/2}. \end{aligned}$$

In cDCC models, S is the variance-covariance matrix of $Q_t^{*1/2} \epsilon_t$ and a (modified) correlation targeting procedure can be led. The cDCC/DCC-GARCH models are well-suited for relatively homogeneous portfolios, for example asset classes from the same geographical area. Actually, the scalar DCC and cDCC specifications provide empirically very close results. Therefore, in our empirical study, we will consider the scalar DCC together with the diagonal QFDCC.

Denoting by S_{++}^n the set of symmetric and positive definite matrices, the vector of parameters for the correlation dynamics in the scalar DCC/cDCC and diagonal QFDCC models are respectively

$$\begin{aligned} \theta_{cor,dcc} &= (S, \alpha, \beta)' \in S_{++}^n \times \mathbb{R}_+^2, \text{ and} \\ \theta_{cor,qfdcc} &= (S, c_{11}, \dots, c_{NN}, a_{11}, \dots, a_{NN}, b_{11}, \dots, b_{NN})' \in S_{++}^n \times \mathbb{R}_+^{3N}. \end{aligned}$$

4 New dynamics for correlations: Vine-GARCH models

4.1 Our specification

In a DCC-type model, one has to rely on intricate normalizations to build sequences of ϵ_t correlation matrices. This makes the interpretation of the (R_t) dynamics not intuitive, because it is deduced from another underlying process (Q_t) . Another drawback of the DCC is the lack of parsimony. In its full form, $O(N^2)$ parameters enter in θ_{cor} , the vector of the parameters

of the correlation process. Hence the Full DCC is intractable since the number of parameters grows rapidly, and creates the same problems as general BEKK models. Most of the time, the DCC is used in its scalar form, but this modeling often fails in capturing fine-tuned and heterogeneous correlation dynamics.

Now, we develop a method which ensures both parsimony and positive definiteness without relying on any normalization. Basically, the idea relies on the modeling of a set of partial correlations, enabling to parameterize any correlation matrix. We use a given regular vine to specify these partial correlations and the one-to-one mapping between these $N(N-1)/2$ partial correlations and the $N(N-1)/2$ “usual” correlations. The former are stacked in a vector P_{C_t} and the latter are the coefficients of R_t . We order partial correlations lexicographically, from the shortest to the longest sets of indices. Then, we propose the following correlation dynamics:

$$\begin{aligned} H_t &= D_t^{1/2} R_t D_t^{1/2}, \\ \Psi(P_{C_t}) &= \Omega + \Xi \Psi(P_{C_{t-1}}) + \Lambda \zeta_{t-1}, \\ R_t &= \text{vech}of(F_{vine}(\Psi^{-1}(P_{C_t}))), \end{aligned}$$

where

- $\text{vech}of(\cdot)$ denotes the operator “devectorization”, that transforms a vector into a symmetric matrix. It is the opposite of the usual operator $\text{vech}(\cdot)$.
- In full generality, Ξ and Λ are $N(N-1)/2 \times N(N-1)/2$ squared matrices of unknown parameters, and Ω is an $N(N-1)/2$ unknown vector. Set the vector of parameters $\theta_{cor} = (\Omega, \Xi, \Lambda)$.
- The vector P_{C_t} is the “partial correlation vector” deduced from a given R-vine structure.
- The vector ζ_t consists of a relevant function of the “innovations”, to update our partial correlations at time t .
- We apply an analytic transformation Ψ to P_{C_t} . For the sake of simplicity, Ψ will be known, even if the methodology can be adapted easily to a parametric function Ψ_θ . To fix the ideas, the multivariate Ψ function will be defined as follows:

$$\begin{aligned} \Psi &:]-1, 1[^{N(N-1)/2} \longrightarrow \mathbb{R}^{N(N-1)/2}, \\ \Psi(P_{C_t}) &= \left(\psi(\rho_{1,2,t}), \dots, \psi(\rho_{N,N-1|L_{N-1,N},t}) \right)', \\ \psi(x) &= \tan(\pi x/2). \end{aligned}$$

The function Ψ twists the univariate dynamics to manage the constraints that partial correlations stay between (-1) and 1 . Alternatively, Ψ could be chosen among the sigmoid functions for instance, for which $\psi(x) = (\exp(\alpha x) - 1) / (\exp(\alpha x) + 1)$, $\alpha \in \mathbb{R}$.

- The function F_{vine} corresponds to the one-to-one mapping from the vector of partial correlations Pc_t to correlations (in R_t) by using the algorithm of Lewandowski, Kurowicka and Joe (2009). It is defined as

$$F_{vine} :]-1, 1[^{N(N-1)/2} \longrightarrow]-1, 1[^{N(N-1)/2},$$

$$F_{vine} (\rho_{1,2,t}, \dots, \rho_{N-1,N|L,t}) = (\rho_{1,2,t}, \dots, \rho_{N-1,N,t})'.$$

Such mapping can be coded easily in the case of C-vines and D-vines. In the general case of R-vines, it is more complex but remains feasible. Note that the function F_{vine} and its inverse are available in the R-package ‘‘Vine-copula’’, under the name ‘‘RVineCor2pcor’’.

To be more specific about ζ_t , for any $L \subset \{1, \dots, N\}$ and $k \notin L$, define $v_{k|L,t}$ by

$$v_{k|L,t} = \frac{\epsilon_{k,t} - \mathbb{E}[\epsilon_{k,t}|\epsilon_{L,t}]}{\sqrt{h_{k|L,t}}},$$

where $\epsilon_{L,t} = (\epsilon_{i,t})_{i \in L}$, and $\mathbb{E}[\epsilon_{k,t}|\epsilon_{L,t}]$ corresponds to the orthogonal projection of the variable $\epsilon_{k,t}$ on the space spanned by the vector $\epsilon_{L,t}$. The variance of the ‘‘residual’’ $\epsilon_{k,t} - \mathbb{E}[\epsilon_{k,t}|\epsilon_{L,t}]$ is denoted by $h_{k|L,t}$. The variables $v_{k|L,t}$ are not really observable, but the next subsection specifies how we can evaluate $\mathbb{E}[\epsilon_{k,t}|\epsilon_{L,t}]$, $h_{k|L,t}$, and we will be able to get $\hat{v}_{k|L,t}$, an approximated value of $v_{k|L,t}$. Then, the $N(N-1)/2$ -sized vector of ‘‘innovations’’ ζ_t stacks the variables $\hat{v}_{i|L,t}\hat{v}_{j|L,t}$, when $(i, j|L)$ is an edge of the underlying vine. The order of these edges in ζ_t will be the same as for Pc_t .

In this paper, to simplify, we will restrict ourselves to a simpler vine-GARCH structure, where there will be no cross-effects between all the individual partial correlation processes. The $N-1$ first elements of Pc_t corresponds to usual correlations, i.e. $\rho_{ij|\emptyset,t} = \rho_{ij,t}$. Each of them is described by the process

$$\psi(\rho_{ij,t}) = \omega_{ij} + \xi_{ij}\psi(\rho_{ij,t-1}) + \lambda_{ij}\hat{v}_{i,t-1}\hat{v}_{j,t-1},$$

with $\hat{v}_{k,t} = \frac{\epsilon_{k,t}}{\sqrt{\hat{h}_{k,t}}}$ are estimated ‘‘degarched’’ residuals. From the N -th on, the elements of Pc_t are true partial correlations for which $L \neq \emptyset$. Their dynamics are given by

$$\psi(\rho_{ij|L,t}) = \omega_{ij|L} + \xi_{ij|L}\psi(\rho_{ij|L,t-1}) + \lambda_{ij|L}\hat{v}_{i|L,t-1}\hat{v}_{j|L,t-1}.$$

Our global process depends on the following set of parameters:

$$\theta_{cor} = \left(\omega_{12}, \dots, \omega_{N-1,N|L_{N-1}}, \xi_{12}, \dots, \xi_{N-1,N|L_{N-1,N}}, \lambda_{12}, \dots, \lambda_{N-1,N|L_{N-1,N}} \right)'$$

$$\in \mathbb{R}^{N(N-1)/2} \times]0, 1[^{N(N-1)/2} \times \mathbb{R}^{N(N-1)/2}.$$

Actually, the structure of the underlying R-vine may be seen as an additional parameter, whose choice is discussed in Subsection 4.3 independently of θ_{cor} . To avoid some explosive behaviors, we impose $\xi_{12}, \dots, \xi_{N-1,N|L_{N-1,N}} \in]0, 1[^{N(N-1)/2}$. Indeed, iterating n times the updating relation of $\psi(\rho_{ij|L,t})$, we get

$$\begin{aligned} \psi(\rho_{ij|L,t}) &= \omega_{ij|L} + \xi_{ij|L} \psi(\rho_{ij|L,t-1}) + \lambda_{ij|L} \hat{v}_{i|L,t-1} \hat{v}_{j|L,t-1} \\ &= \omega_{ij|L} \sum_{k=0}^{n-1} \xi_{ij|L}^k + \xi_{ij|L}^n \psi(\rho_{ij|L,t-n}) + \lambda_{ij|L} \sum_{k=0}^{n-1} \xi_{ij|L}^k \hat{v}_{i|L,t-k-1} \hat{v}_{j|L,t-k-1}. \end{aligned}$$

Since n may be arbitrarily large, a necessary condition for convergence is clearly $\xi_{ij|L} < 1$.

In full generality, this simplified version of the vine-GARCH model still encompasses $3N(N-1)/2$ parameters. However, this approach can become more parsimonious easily and would provide a nice alternative to full DCC-GARCH models. Indeed, when the p th-VF property applies¹, one can set constraints to any level of the tree (say p), and choose zero partial correlations at and after the p -th tree in the underlying vine. We guess this should not modify too strongly the (true) correlation dynamics, at least when p is not too small. This is due to the fact that partial correlations with non-empty conditioning subsets are correlations between residuals. In practice it is likely that these residuals behave more and more as white noise when the number of conditioning variables increases. Indeed, when the cardinality of the conditioning subsets L goes up, we are building (standardized) residuals $v_{k|L,t}$ on larger and larger subspaces.

For instance, for a C-vine, the first conditioning subset will be the asset that is the most correlated with most others (the S&P500, for instance), i.e. the main factor $X_{a_1,t}$ of the model. After controlling for the main factor, partial correlations of the type $\rho_{i,j|a_1}$ focus on the dependence between projections of random vectors on the orthogonal space of X_{a_1} . Knowing X_{a_1} , the second main factor is X_{a_2} , and we expect that the partial correlations $\rho_{i,j|a_1,a_2}$ will be smaller than the ones $\rho_{i',j'|a_1}$, i.e. partial correlations associated to the edges of the preceding tree. This is not a mathematical result, but empirical intuition that has to be validated. Indeed, in a lot of situations in practice, some variables appear as natural factors (“drivers”) for the others. In an equity portfolio, one of the main stock indices may play this role. In the fixed income universe, some interest rates (or tenor of the IR curve) are more important than others for dealers because they correspond to the maturity of existing futures. Etc.

In every case, by canceling partial correlations after the step p in a vine, we get a particular model with less parameters than in the full vine-GARCH specification. If the vine is p -VF, whatever the chosen structure of the vine is after level p , the reconstruction formulas (1) provide the same correlation matrices R_t . This is a nice theoretical property. A slightly

¹and it is the case if we restrict ourselves to C-vines or D-vines (see Subsection 2.4)

different simplification of our vine-GARCH models would be to assume constant (non zero) partial correlations after some level (say p) in the vine. But in this case, we cannot insure a similar p -VF property.

4.2 Determination of ζ_t

At time t , the vector ζ_t is a key information as it drives the shocks on the correlation processes. Here, we propose two ways of evaluating $\hat{v}_{k|L,t}$, the key components to calculate ζ_t .

The first method is based on the linear regression of ϵ_t on $\epsilon_{L,t}$:

$$\epsilon_{k,t} = \alpha_{k|L} + \beta'_{k|L} \epsilon_{L,t} + \xi_{L,t}, \quad \mathbb{E}[\xi_{L,t} | \epsilon_{L,t}] = 0,$$

where $\alpha_{k|L}$ is a scalar and $\beta_{k|L}$ is a $\#L$ -dimensional vector. Then, we approximate $\epsilon_t - \mathbb{E}[\epsilon_t | \epsilon_{L,t}]$ by $\epsilon_t - \hat{\alpha}_{k|L} - \hat{\beta}'_{k|L} \epsilon_{L,t}$. Moreover, an empirical “rolling-window” estimator of $h_{k|L,t}$ can be defined by

$$\hat{h}_{k|L,t} := \frac{1}{T} \sum_{i=1}^T \left(\epsilon_{k,t-i} - \left(\hat{\alpha}_{k|L} + \hat{\beta}'_{k|L} \epsilon_{L,t-i} \right) \right)^2,$$

for some windows size T . This approach may be termed “non parametric” in the sense that it does not rely on any hypothesis about the conditional distribution of ϵ_t . Then, this first method provides

$$\hat{v}_{k|L,t} := \frac{\epsilon_{k,t} - \hat{\alpha}_{k|L} - \hat{\beta}'_{k|L} \epsilon_{L,t}}{\sqrt{\hat{h}_{k|L,t}}}.$$

The second method is based on the theoretical distribution of the residuals $\epsilon_t | \mathcal{F}_{t-1}$, that is unknown at this stage. In our case, we have assumed this distribution is Gaussian. Recall that, if a vector $(X, Y)'$ is Gaussian, then

$$\begin{aligned} X|Y &\sim \mathcal{N}(E_{X|Y}, \Sigma_{X|Y}), \quad \text{where} \\ E_{X|Y} &= \mathbb{E}[X] + \text{cov}(X, Y) \text{Var}(Y)^{-1} (Y - \mathbb{E}[Y]), \\ \Sigma_{X|Y} &= \text{Var}(X) - \text{cov}(X, Y) \text{Var}(Y)^{-1} \text{cov}(X, Y)'. \end{aligned}$$

Through these formulas and under our model assumptions, we can calculate

$$v_{k|L,t} = \frac{\epsilon_{k,t} - \mathbb{E}[\epsilon_{k,t} | \epsilon_{L,t}]}{\sqrt{h_{k|L,t}}}.$$

Indeed, as $(\epsilon_{k,t}, \epsilon_{L,t})'$ is supposed to be Gaussian and centered, we obtain

$$\mathbb{E}[\epsilon_{k,t} | \epsilon_{L,t}] = \text{cov}_{t-1}(\epsilon_{k,t}, \epsilon_{L,t}) \text{Var}_{t-1}(\epsilon_{L,t})^{-1} \epsilon_{L,t}, \quad \text{and}$$

$$\begin{aligned}
h_{k|L,t} &= \text{Var}_{t-1}(\epsilon_{k,t} - \mathbb{E}[\epsilon_{k,t}|\epsilon_{L,t}]) \\
&= \text{Var}_{t-1}(\epsilon_{k,t}) - \text{cov}_{t-1}(\epsilon_{k,t}, \epsilon_{L,t}) \text{Var}_{t-1}(\epsilon_{L,t})^{-1} \text{cov}_{t-1}(\epsilon_{L,t}, \epsilon_{k,t}).
\end{aligned}$$

To give an example, suppose $L = (j)$ and $\epsilon_{L,t} = \epsilon_{j,t}$. Then, we have

$$\mathbb{E}[\epsilon_{k,t}|\epsilon_{j,t}] = \rho_{kj,t} h_{kk,t} \epsilon_{j,t} / h_{jj,t}, \text{ and}$$

$$\begin{aligned}
\text{Var}_{t-1}(\epsilon_{k,t} - \mathbb{E}[\epsilon_{k,t}|\epsilon_{L,t}]) &= h_{kk,t}^2 - \rho_{kj,t} h_{kk,t} h_{jj,t} (h_{jj,t}^2)^{-1} \rho_{kj,t} h_{kk,t} h_{jj,t} \\
&= h_{kk,t}^2 (1 - \rho_{kj,t}^2).
\end{aligned}$$

Then,

$$\hat{v}_{k|L,t} = \frac{\epsilon_{k,t} - \rho_{kj,t} \hat{h}_{kk,t} \epsilon_{j,t} / \hat{h}_{jj,t}}{\hat{h}_{kk,t} \sqrt{1 - \rho_{kj,t}^2}},$$

where $\hat{h}_{kk,t}$ is given by the k -th univariate GARCH process.

In this paper, the second method of calculation will be used to get (ζ_t) .

4.3 Vine selection

The methodology above can be applied to any R-vine on N elements. Nonetheless, selecting a convenient R-vine may be useful to describe the dependence among the variables in a parsimonious and meaningful way. In particular, this would allow for the truncation of a given R-vine (as explained in Subsection 2.4), once some important factors have been found in the first trees.

To do so with a C-vine, we can follow the sequential method developed by Dissmann et al. (2012). This method consists in starting from the first tree, computing the Kendall's tau of all its edges, and selecting the variable which provides the highest degree of dependence with the other ones. Conditionally on this variable and in the second tree, we compute a Kendall's tau per edge, but conditional on the variable chosen on the first tree. We apply the same selection criteria to choose the right variable and proceed on the next tree, until the last tree. The Kendall's tau is used as a dependence measure because it is free of any marginal distribution and because it can be easily calculated, but other choices exist. All the details concerning these steps together with the empirical Kendall's tau computation are reported in Appendix B.

The latter selection procedure is “bottom-up”. Alternatives exist in the general case of R-vines, in particular the “top-down” procedure of Kurowicka (2011).

5 Statistical inference: Quasi Maximum Likelihood estimation

We now turn to the statistical estimation of our previous vine-GARCH model. To do so, some estimators are obtained by maximizing a likelihood function that does not correspond to the true Data Generating Process necessarily. When such a procedure is applied to the estimation of the parameters of the second order moments, relying on Gouriéroux, Monfort and Trognon (1984) and Bollerslev and Wooldridge (1994) for instance, this so-called Quasi-Maximum Likelihood Estimator (QMLE) is consistent and asymptotically normal, under appropriate assumptions. Our specification in (5) corresponds to a Gaussian QMLE and we apply a two-step estimation method, that is usual in this stream of the literature. In other words, we shall assume that (η_t) is a Gaussian white noise for inference purpose only, but the “true” underlying distribution of η_t may be different ².

5.1 The QML estimator

We suppose that $\epsilon_t|\mathcal{F}_{t-1}$ follows a multivariate normal distribution. Its density is given by

$$\begin{aligned} f(\epsilon_t|\mathcal{F}_{t-1}, \theta) &= \frac{1}{(2\pi)^{N/2} |H_t|^{1/2}} \exp \left\{ -\frac{1}{2} \epsilon_t' H_t^{-1} \epsilon_t \right\} \\ &= \exp \left\{ -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log(|H_t|) - \frac{1}{2} \epsilon_t' H_t^{-1} \epsilon_t \right\}. \end{aligned}$$

This likelihood f is quadratic in ϵ_t and clearly belongs to the quadratic exponential family (see Gouriéroux et al. 1984). Using the joint independence of the η_t and developing H_t as $D_t^{1/2} R_t D_t^{1/2}$, the likelihood function becomes

$$\begin{aligned} L_T(\theta; \epsilon) &= \prod_{t=1}^T \exp \left\{ -\frac{N}{2} \log(2\pi) - \frac{1}{2} \log(|H_t|) - \frac{1}{2} \epsilon_t' H_t^{-1} \epsilon_t \right\} \\ &= \prod_{t=1}^T \exp \left\{ -\frac{1}{2} \left(N \log(2\pi) + \log(|D_t^{1/2} R_t D_t^{1/2}|) + \epsilon_t' D_t^{-1/2} R_t^{-1} D_t^{-1/2} \right) \right\} \\ &= \prod_{t=1}^T \exp \left\{ -\frac{1}{2} \left(N \log(2\pi) + \log(|D_t|) + \log(|R_t|) + u_t' R_t^{-1} u_t \right) \right\} \\ &= \prod_{t=1}^T \exp \left\{ -\frac{1}{2} \left(N \log(2\pi) + \log(|D_t|) + \epsilon_t' D_t^{-1} \epsilon_t - u_t' u_t + \log(|R_t|) + u_t' R_t^{-1} u_t \right) \right\}, \end{aligned}$$

²Note that it can be estimated empirically a posteriori from a sample of residuals $R_t(\hat{\theta})^{-1/2} \epsilon_t$

where $D_t^{1/2} = \text{diag}(\sqrt{h_{11,t}}, \dots, \sqrt{h_{NN,t}})$, and $u_t = (\epsilon_{1,t}/\sqrt{h_{11,t}}, \dots, \epsilon_{N,t}/\sqrt{h_{NN,t}})'$ is the vector of GARCH standardized residuals. Thus, the log-likelihood function can be expressed as

$$\begin{aligned} \ell(\theta; \epsilon) &= \log \left(\prod_{t=1}^T f(\epsilon_t | \mathcal{F}_{t-1}, \theta) \right) \\ &= \sum_{t=1}^T \left[-\frac{N}{2} \log(2\pi) - \frac{1}{2} \Delta(\epsilon_t; \theta_{vol}) - \frac{1}{2} \Upsilon(\epsilon_t; \theta_{vol}, \theta_{cor}) + \frac{1}{2} u_t' u_t \right], \end{aligned} \quad (8)$$

with

$$\begin{aligned} \Delta(\epsilon_t; \theta_{vol}) &= \log(|D_t|) + \epsilon_t' D_t^{-1} \epsilon_t, \text{ and} \\ \Upsilon(\epsilon_t; \theta_{vol}, \theta_{cor}) &= \log(|R_t|) + u_t' R_t^{-1} u_t. \end{aligned}$$

$\Delta(\cdot)$ and $\Upsilon(\cdot)$ are quadratic functions of ϵ_t . Only θ_{vol} enters in $\Delta(\cdot)$ whereas both θ_{vol} and θ_{cor} enter in $\Upsilon(\cdot)$. Hence $\Delta(\cdot)$ is the volatility part and conditional on θ_{vol} , $\Upsilon(\cdot)$ is the correlation part. This is a classic framework of a two-step Quasi Maximum Likelihood in the sense that the problem can be split into two steps. In our simplified case, the volatility part of the log-likelihood function corresponds to the sum of log-likelihood functions of N univariate variance models that can be estimated independently:

$$QL1_T(\theta_{vol}; \epsilon) = -\frac{1}{2} \sum_{i=1}^N \sum_{t=1}^T \left[\log(h_{iit}) + \frac{\epsilon_{i,t}^2}{h_{iit}} \right].$$

Given θ_{vol} , and under appropriate regularity conditions of $\Upsilon(\cdot)$, a consistent, but inefficient, estimator of θ_{cor} can be obtained by maximizing:

$$QL2_T(\theta_{cor}; \theta_{vol}, \epsilon) = -\frac{1}{2} \sum_{t=1}^T [\log(|R_t|) + u_t' R_t^{-1} u_t], \quad (9)$$

with $u_t = D_t^{-1/2} \epsilon_t$. Since maximizing ℓ is equivalent to maximizing the sum of $QL1_T(\theta_{vol}; \epsilon)$ and $QL2_T(\theta_{cor}; \theta_{vol}, \epsilon)$, the two-step approach for maximizing (8) can be led consequently. The first step consists in finding

$$\hat{\theta}_{vol} = \arg \max_{\theta_{vol}} QL1_T(\theta_{vol}; \epsilon). \quad (10)$$

Conditioning on $\hat{\theta}_{vol}$, the only portion of the ℓ that will influence the parameter selection is $\log(|R_t|) + u_t' R_t^{-1} u_t$. So we exclude constant terms and the second step of the QML estimation, conditional on $\hat{\theta}_{vol}$, consists in optimizing

$$\hat{\theta}_{cor} = \arg \max_{\theta_{cor}} QL2_T(\theta_{cor}; \hat{\theta}_{vol}, \epsilon). \quad (11)$$

Note that the parameter $\hat{\theta}_{cor}$ determines the volatility processes $(h_{ii,t})$ and then the (now estimated) residuals u_t .

Following Theorem 1 of Engle and Sheppard (2001), under appropriate conditions, especially the existence of stationary and ergodic solutions, the two-step QML estimator is consistent. In particular, $\hat{\theta}_{cor} \xrightarrow{P} \theta_{0,cor}$. As for the asymptotic behavior, we rely on Bollerssev and Wooldridge (1994) and Engle and Sheppard's results (2001). According to Theorem 2 of Engle and Sheppard, the two-step QML estimator follows asymptotically a normal distribution with Bollerssev-Wooldridge type variance covariance matrix:

$$\sqrt{T} \left(\left(\hat{\theta}_{vol}, \hat{\theta}_{cor} \right)' - (\theta_{0,vol}, \theta_{0,cor})' \right) \xrightarrow{d} N(0, A_0^{-1} B_0 A_0^{-1'}), \quad (12)$$

with:

$$A_0 = \frac{1}{2} \begin{bmatrix} \partial_{\theta_{vol}\theta_{vol}}^2 \Delta(\epsilon_t; \theta_{vol}) & 0 \\ \partial_{\theta_{vol}\theta_{cor}}^2 \Upsilon(\epsilon_t; \theta_{vol}, \theta_{cor}) & \partial_{\theta_{cor}\theta_{cor}}^2 \Upsilon(\epsilon_t; \theta_{vol}, \theta_{cor}) \end{bmatrix} := \begin{bmatrix} A_{11} & 0 \\ A_{12} & A_{22} \end{bmatrix},$$

and

$$B_0 = \frac{1}{4T} Var \left[\sum_{t=1}^T \left\{ \partial_{\theta_{vol}} \Delta(\epsilon_t; \theta_{vol}), \partial_{\theta_{cor}} \Upsilon(\epsilon_t; \theta_{vol}, \theta_{cor}) \right\} \right] := \begin{bmatrix} B_{11} & B_{12} \\ B_{12} & B_{22} \end{bmatrix}.$$

5.2 Estimation strategy

Let us specify the practical way of optimizing the previous QML criteria.

First, we build an estimator of $\theta_{vol} = (\varsigma_1, \kappa_1, \tau_1, \dots, \varsigma_N, \kappa_N, \tau_N)$ as in (10), according to the QML method. Since we do as if the law of ϵ_t knowing \mathcal{F}_{t-1} is Gaussian, for every $i = 1, \dots, N$, the first order orthogonality conditions are given by

$$\sum_{t=1}^T \left(\frac{1}{h_{ii,t}} - \frac{\epsilon_{i,t}^2}{h_{ii,t}^2} \right) \partial_{\theta_{vol}} (\varsigma_i + \kappa_i \epsilon_{i,t-1}^2 + \tau_i h_{ii,t-1}) = 0.$$

Under appropriate constraints, the resolution of the latter system of equations provides an estimate of θ_{vol} . As usual for GARCH(1,1) models, we impose the positivity of the parameters and a sufficient condition to get stationarity solutions is $\kappa_i + \tau_i \leq 1$ for all i . A numerical procedure is required for solving this system. To do so, we use the Newton-Raphson method. Finally, the variance of those estimators is estimated by taking the empirical counterpart of the Fisher matrix, corrected according to Bollerssev and Wooldridge. All the details related to those technics are given in Section A in the appendix.

Second, conditional on the first step estimation, we focus on the optimization (11). We

build an estimator of θ_{cor} according to the QML method. Due to the nullity of the score in expectation, $\mathbb{E}[\partial_{\theta_{cor}} \Upsilon(\epsilon_t; \theta_{vol}, \theta_{cor})] = 0$. Therefore, the empirical counterpart of this equation defines the estimator $\hat{\theta}_{cor}$ as the root of

$$\sum_{t=1}^T \partial_{\theta_{cor}} \Upsilon(\epsilon_t; \hat{\theta}_{vol}, \theta_{cor}) = 0.$$

The underlying processes (R_t) induces tricky computations of scores and Hessians, for both DCC and vine Garch dynamics. In this paper, we consider two strategies depending on the dimensionality of the problem. The details are given in the appendix, Subsection A.2.

5.2.1 Sequential Quadratic Programming

The estimation strategy depends on the complexity of each parametrization. In this study, DCC specifications are not highly parameterized: the scalar DCC (resp. diagonal QFDCC) requires the estimation of 3 (resp. $3N$) parameters, after correlation targeting. Consequently, the Sequential Quadratic Programming method is implemented for these DCC dynamics since it is well-suited for constrained optimization with a “reasonable” number of parameters.

As the general DCC model, the vine GARCH specification may suffer from the curse of dimensionality. However, it is possible to weaken this problem strongly by using some nice theoretical properties of C-vines. Actually, a vine GARCH model based on an underlying C-Vine may be estimated by solving $N \times (N - 1)/2$ bivariate optimization programs, instead of through a single “brute-force” optimization stage. This is due to the fact that, in a C-vine, each partial correlation on tree T_i can be updated easily knowing the partial correlations on tree T_{i-1} . This simple property allows for proceeding sequentially. In other words, we estimate successively $N \times (N - 1)/2$ bivariate dynamic models that are related to the dynamics of $(\epsilon_{i,t}, \epsilon_{j,t})$ when the couples (i, j) describe the conditioned subsets of all the nodes in the underlying C-vine, starting from the bottom tree.

To be more explicit, assuming that 1 is the root in the first tree, consider the $N - 1$ first partial correlation dynamics given by

$$\psi(\rho_{i1,t}) = \omega_{i1} + \xi_{i1} \psi(\rho_{i1,t-1}) + \lambda_{i1} \hat{v}_{i,t-1} \hat{v}_{1,t-1}, \quad j = 2, \dots, N, \quad (13)$$

where $\hat{v}_{k,t} = \frac{\epsilon_{k,t}}{\sqrt{\hat{h}_{k,t}}}$ are “degarched” residuals. The latter quantities depend on the estimated volatility processes and the observations. Consequently, we can estimate the parameters of any equation (13) *independently* by maximizing the next $N - 1$ objective functions over there

corresponding correlation spaces:

$$\left\{ \begin{array}{l} QL2_T^{12} \left(\theta_{cor,12}; \hat{\theta}_{vol}, \epsilon \right) = -\frac{1}{2} \sum_{t=1}^T \left[\log (|R_{(12),t}|) + u'_{(12),t} (R_{(12),t})^{-1} u_{(12),t} \right], \\ QL2_T^{13} \left(\theta_{cor,13}; \hat{\theta}_{vol}, \epsilon \right) = -\frac{1}{2} \sum_{t=1}^T \left[\log (|R_{(13),t}|) + u'_{(13),t} (R_{(13),t})^{-1} u_{(13),t} \right], \\ \dots \\ QL2_T^{1N} \left(\theta_{cor,1N}; \hat{\theta}_{vol}, \epsilon \right) = -\frac{1}{2} \sum_{t=1}^T \left[\log (|R_{(1N),t}|) + u'_{(1N),t} (R_{(1N),t})^{-1} u_{(1N),t} \right]. \end{array} \right.$$

Here, for all $j = 2, \dots, N$, $u_{(1j),t} = [\hat{v}_{1,t}, \hat{v}_{j,t}]'$; $R_{(1j),t}$ are the 2×2 correlation matrices and $\theta_{cor,1j} = (\omega_{1j}, \xi_{1j}, \lambda_{ij})$ are the parameters that are associated to the bivariate process $(\epsilon_{1,t}, \epsilon_{j,t})$.

Now suppose 2 is the root of the next tree T_2 , when conditioning by 1. There are $N - 2$ dynamic partial correlations in T_2 . Their processes are described as follows:

$$\forall j = 3, \dots, N, \psi(\rho_{2j|1,t}) = \omega_{2j|1} + \xi_{2j|1} \psi(\rho_{2j|1,t-1}) + \lambda_{2j|1} \hat{v}_{2|1,t-1} \hat{v}_{j|1,t-1}.$$

For inference purpose, we assumed $\epsilon_t | \mathcal{F}_{t-1} \sim \mathcal{N}(0, H_t)$. Consequently, we use the second method described in 4.2 to compute $\hat{v}_{2|1,t-1}$ and $\hat{v}_{j|1,t-1}$. Under the Gaussian assumption, partial correlations and conditional correlations are similar. Therefore, for any $j = 3, \dots, N$, the joint distribution of $(u_{2,t}, u_{j,t})$ conditional on $u_{1,t}$ is normal, with covariance/correlation $\rho_{2j|1,t}$. The latter quantity is an explicit function of the correlations $\rho_{12,t}$ and $\rho_{1j,t}$ that have been estimated at the previous step on T_1 . Hence, we can also estimate the previous partial correlations dynamics by maximizing $N - 2$ objective functions independently over each correlation parameter space of tree T_2 , conditional on the estimated correlations on T_1 :

$$\left\{ \begin{array}{l} QL2_T^{23|1} \left(\theta_{cor,23|1}; \hat{\theta}_{vol}, \epsilon, \hat{\rho}_{12}, \hat{\rho}_{13} \right) = -\frac{1}{2} \sum_{t=1}^T \left[\log (|R_{(23),t}|) + u'_{(23),t} (R_{(23),t})^{-1} u_{(23),t} \right], \\ QL2_T^{24|1} \left(\theta_{cor,24|1}; \hat{\theta}_{vol}, \epsilon, \hat{\rho}_{12}, \hat{\rho}_{14} \right) = -\frac{1}{2} \sum_{t=1}^T \left[\log (|R_{(24),t}|) + u'_{(24),t} (R_{(24),t})^{-1} u_{(24),t} \right], \\ \dots \\ QL2_T^{2N|1} \left(\theta_{cor,2N|1}; \hat{\theta}_{vol}, \epsilon, \hat{\rho}_{12}, \hat{\rho}_{1N} \right) = -\frac{1}{2} \sum_{t=1}^T \left[\log (|R_{(2N),t}|) + u'_{(2N),t} (R_{(2N),t})^{-1} u_{(2N),t} \right]. \end{array} \right.$$

Here, for every $j = 2, \dots, N$, $R_{(2j),t}$ is the 2×2 correlation matrix with extra-diagonal elements $\rho_{2j,t}$. The latter coefficient is computed from the estimated dynamic partial correlations $\hat{\rho}_{2j|1,t}$ and correlations $\hat{\rho}_{1j|1,t}$. Obviously, $u_{(j,k),t} = [\hat{v}_{j,t}, \hat{v}_{k,t}]'$.

We apply the same reasoning for the next trees in the C-vine. There are $N - 3$ objective functions to be maximized on tree T_3 , $N - 4$ on tree T_4 , etc, until tree T_{N-1} where only one objective function needs to be maximized. The iterative algorithm is based on the idea of a

multi-step estimation procedure relying on the vine structure. The estimation of the partial correlation process of a tree T_i only depends on a subset of partial correlations associated to the nodes of T_{i-1} , invoking the recursive formula (1). Thus, denoting $\theta_{cor,ij|L} = (\omega_{ij|L}, \xi_{ij|L}, \lambda_{ij|L})$ a subvector of θ_{cor} , our iterative algorithm can be summarized as follows, assuming 1 as the root in T_1 , 2 the root on tree T_2 , and the like until tree T_{N-1} and denoting the conditioning set of T_i as $L_i = \{1, \dots, i-1\}$, $i = 2, \dots, N-1$:

$$\hat{\theta}_{cor,ij|L_i} = \arg \max_{\theta_{cor,ij|L_i}} QL2_T^{ij|L_i} \left(\theta_{cor,ij|L_i}; \hat{\theta}_{vol}, \epsilon, \hat{\rho}_{i-1,i|L_{i-1}}, \hat{\rho}_{i-1,j|L_{i-1}}, \hat{\rho}_{ij|L_i} \right),$$

for every i and j in $\{1, \dots, N\}$, $i < j$.

We denote this strategy *C-Vine iterative process*, which is particularly effective when N becomes large (say from 8 assets). At each node on a specific level, only 3 parameters need to be estimated. Consequently, we also use the Sequential Quadratic Programming method when estimating the C-Vine iterative process.

5.2.2 Simulated Annealing

A drawback of the latter C-Vine iterative process is the diffusion of estimation errors from one partial correlation level to the next one. It is still possible to estimate the vine-GARCH at once for reasonable portfolio sizes ($N \leq 8$) to avoid this iterative method. But since F_{vine} induces a high level of nonlinearity, the calculation of the Hessian may be painful. Applying numerical derivatives, we can use some algorithms that require only first derivatives such as the Bernd-Hall-Hall-Hausmann algorithm. But their performances may decrease dramatically when ℓ is highly nonlinear. Besides, if ℓ admits several local maxima, the Newton-Raphson or the Bernd-Hall-Hall-Hausmann methods depend strongly on the initial parameter values before running the iterations. Therefore, the nonlinearity and the instability of the likelihood function in the vine-GARCH case require another approach to maximize (11). Here, we use a stochastic algorithm, the simulated annealing. It enables to simulate the objective function from some given values of θ_{cor} . The main idea is that a change of scale, called temperature, enables faster moves on the surface of $QL2_T$ to be maximized. By rescaling partially, this method prevents from falling in local maxima. See details in Subsection A.2.2 in the appendix.

Note that the simulated annealing algorithm can also be used when estimating model through the previous C-Vine iterative methodology. However in this case, the Sequential Quadratic Programming is a lot quicker, which is the reason we used this method in the simulation study.

6 Empirical applications

6.1 Simulation study

We consider as a data generating process (DGP) a multivariate series (ϵ_t) of size $N = 6$ that is simulated from a centered normal law. Its conditional variance-covariance matrices are deduced from an MGARCH form $H_t = D_t^{1/2} R_t D_t^{1/2}$. To generate N univariate variance processes along (7), we choose the $3N$ parameters as random such that $\varsigma \sim U(0.00001, 0.00009)$, $\kappa \sim U(0.01, 0.15)$ and $\tau \sim U(0.95, 0.85)$, under the stationarity constraint $\kappa + \tau < 1$. As for the correlation dynamics, we first choose randomly $N(N - 1)/2$ deterministic processes among the cosinus, sinus, modulo and constant functions, and then generate the patterns

$$\begin{cases} a_1 + a_2 \cos(2\pi t/\alpha), \\ b_1 + b_2 \sin(2\pi t/\beta), \\ c_1 + c_2 \text{mod}(t/\mu), \\ d_1 + d_2 \text{const}, \end{cases}$$

for every $t = 1, \dots, T$, where $a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2$ are chosen randomly following independent $\mathcal{U}(-0.4, 0.4)$ and α, β, μ are randomly (equally) selected among $\{200, 500, 1000, 1500, 2000\}$. Then, for each t , we build a lower triangular matrix K_t with ones on the main diagonal, which is the Cholesky decomposition of the symmetric and positive definite matrix $C_t = K_t K_t'$. Finally, we generate $R_t = C_t^{\star-1/2} C_t C_t^{\star-1/2}$, as for the normalization step of a DCC. Those processes allow for rapid, gradual changes or constant correlation patterns, and they do not depend on a specific statistical model. Initializing each of the GARCH processes and given ϵ_1 , we simulate the successive values of a MGARCH process with conditional variance-covariance matrices (H_t) , $H_t = D_t^{1/2} R_t D_t^{1/2}$. We do this iterative procedure for $T = 10000$ and we consider 300 different correlation matrix patterns. Once a series is simulated, we estimate the model under different model assumptions: a C-Vine-GARCH, a diagonal QFDCC and a scalar DCC models. As a benchmark, we also compute a rolling-window correlation matrix with a rolling-window of size 200 observations. The estimated parameters allow the calculation of successive correlation matrices, which are here \hat{R}_t^{vine} (C-vine-GARCH), \hat{R}_t^{qfdcc} (QFDCC model), \hat{R}_t^{dcc} (DCC model), and \hat{R}_t^{rw} (rolling-window) correlations. For the vine-Garch, we consider a constrained case too, where the partial correlations of the last two trees are constrained to their unconditional partial correlations values, as estimated over the whole sample³. We denote by $\hat{R}_t^{\text{vine}\star}$ the correlation matrices obtained with the constrained version of the C-vine. Both vine specifications are estimated by the C-Vine iterative process. The first level of the C-vine has been chosen following the procedure 4.3.

We are interested in comparing the true correlation process and the estimated correlation

³Alternatively, we could set zero partial correlations for these two last trees of the C-vine. The results would be comparable.

processes through the aforementioned models. To do so, we specify a matrix distance, namely the Frobenius norm, defined as

$$\|A - B\|_F = \sqrt{\text{trace}((A - B)'(A - B))}.$$

We can compute the previous norm for each t and for $A = R_t$ and

$$B \in \{\hat{R}_t^{dcc}, \hat{R}_t^{qfdcc}, \hat{R}_t^{rw}, \hat{R}_t^{vine}, \hat{R}_t^{vine*}\}.$$

We take the average of those quantities over $T = 10000$ periods of time. Since we repeat this experiment 300 times, this provides an average gap for all those simulations. Table 1 reports these results.

Table 1: Simulation study: average distance between true and estimated correlation matrices

	$B = \hat{R}_t^{dcc}$	$B = \hat{R}_t^{qfdcc}$	$B = \hat{R}_t^{rw}$	$B = \hat{R}_t^{vine}$	$B = \hat{R}_t^{vine*}$
$\ R_t - B\ _F$	0.4995	0.4791	0.5275	0.3906	0.4137

The C-Vine model clearly outperforms the other specifications. The DCC displays a significant gap, which highlights that it is too restrictive to capture complex dynamics with only two parameters. As for the rolling-window correlation, the result emphasizes this empirical measure should be taken with great care. The rolling nature of the samples makes the rolling-window correlation very low to react to a rapid correlation fluctuations. The diagonal QFDCC displays better performances compared with the scalar DCC, but it is still significantly outperformed by the C-Vine model. Finally, the constrained vine-Garch still challenges the DCC-type dynamics. Therefore this justifies the use of constrained C-Vine dynamics, allowing for parsimony.

6.2 Application to real portfolios

In this section, we estimate by Quasi-Maximum Likelihood the DCC-GARCH and vine-GARCH models for two portfolios composed of daily series of log-returns related to the twenty-three country indices that compose the Morgan Stanley Capital International (MSCI) Developed Markets. For the first portfolio, denoted portfolio I, we consider Germany, Italy, France, the Netherlands and the United Kingdom. Portfolio II is more diversified geographically because it is composed of Germany, the United-States, Greece, Italy, Japan and Australia. For both portfolios, the samples start in January 1999 and end in August 2014, which amounts to 3669 observations. Table 2 gathers some descriptive statistics about the studied sample. Means, standard deviations, minima and maxima are in percentage terms. Greece and Italy are the only countries that have non positive mean. Importantly, these time series are skewed

and have a significant excess kurtosis, emphasizing non-normality, which is confirmed by the Jarque-Berra tests that reject the normality hypothesis.

Table 2: Descriptive Statistics of the MSCI Log Returns.

Assets	Mean	Std. Dev.	Min	Max	Skewness	Kurtosis	JB stat.
Australia	0.039	1.61	-15.97	16.30	-0.674	14.79	2314***
France	0.014	1.69	-11.57	14.10	0.022	9.10	5690***
Germany	0.014	1.75	-10.77	14.23	-0.031	7.93	3723***
Greece	-0.037	2.37	-34.06	31.62	-0.128	25.90	80167***
Italy	-0.005	1.73	-11.11	15.70	0.012	10.09	7685***
Japan	0.008	1.48	-9.51	10.44	-0.248	6.55	1964***
Netherlands	0.011	1.61	-11.51	14.33	-0.057	9.66	6777***
United-Kingdom	0.014	1.46	-10.43	13.91	-0.039	12.59	1406***
United-States	0.012	1.32	-9.51	10.34	-0.230	10.15	7857***

Significance level: * for 0.05, ** for 0.01, *** for 0.001

Let us notice that we have demeaned the time series by estimating $\mathbb{E}_t[r_t] = \mu_t(\theta)$ by OLS. Then we estimate multivariate conditional variance processes on $\epsilon_t = r_t - \mu_t$. The GARCH(1,1) specification was chosen a priori for modeling the marginal dynamics without a strict selection process. Indeed, this is by far the reference model used in the literature. The estimation results are reported in Table 3.

Table 3: GARCH(1,1) Models estimated by QML for 11 stock indices. The Bollershev-Wooldridge standard deviations are in parentheses.

Asset	ς	κ	τ
Australia	0.657e-005 (0.114e-005)	0.124 (0.014)	0.846 (0.011)
France	0.388e-005 (0.076e-005)	0.111 (0.009)	0.876 (0.008)
Germany	0.368e-005 (0.080e-005)	0.100 (0.011)	0.889 (0.010)
Greece	0.191e-005 (0.147e-005)	0.090 (0.010)	0.917 (0.015)
Italy	0.235e-005 (0.052e-005)	0.113 (0.010)	0.883 (0.008)
Japan	0.997e-005 (0.157e-005)	0.103 (0.012)	0.849 (0.013)
Netherlands	0.363e-005 (0.069e-005)	0.110 (0.010)	0.876 (0.009)
United-Kingdom	0.338e-005 (0.067e-005)	0.115 (0.011)	0.868 (0.009)
United-States	0.223e-005 (0.056e-005)	0.102 (0.010)	0.884 (0.008)

We now turn to the second step QML-estimation of the conditional correlation dynamics, knowing the GARCH(1,1) estimates, for portfolios I and II. We index each country with a

number, which corresponds to the variable selected as the root on each tree according to the Kendall’s tau selection procedure (see Subsection 4.3). Portfolio I is composed of European stocks. Therefore, it can be considered as relatively homogenous, including the main countries of the Eurozone. The selecting procedure forms the following order: Germany (1), United-Kingdom (2), Italy (3), France (4) and Netherlands (5). In this case, Germany (1) is the root of the first C-Vine tree. That means we consider the partial correlation of two countries given Germany on Tree 2. Then, on Tree 3, the conditioning subset is Germany (1) and United-Kingdom (2). Finally, Italy is added in the conditioning set on the next tree. As for portfolio II, which can be considered as a heterogenous portfolio, its composition is given as follows: Germany (1), Greece (2), United-States (3), Italy (4), Japan (5) and Australia (6).

For the C-Vine-GARCH model, we consider two cases. The first is the usual unconstrained C-Vine tree. The second one is a constrained version of the previous one, where the partial correlations of the two last trees are fixed. Therefore, in portfolio I, $\rho_{45|123}$, $\rho_{35|12}$ and $\rho_{34|12}$ are set to their unconditional values that have been estimated over the whole sample. Thus the size of the parameter space is reduced by 9 parameters for both portfolios. For both portfolios, the models are estimated by simulated annealing. Table 4 reports the estimation results of the vine-GARCH model for the unconstrained case. The results for the constrained case are very close to those of the unconstrained case and are reported in Section 10 in the appendix, Tables 12 and 13. Table 6 (resp. Table 5) reports the estimation results of the scalar DCC (resp. diagonal QFDCC).

Table 4: C-Vine-GARCH Model estimated by QML for portfolio I: Germany (1), United-Kingdom (2), Italy (3), France (4), Netherlands (5). The Bollershev-Wooldridge standard deviations are in parentheses.

Ω	Estimate (Std Err)	Ξ	Estimate (Std Err)	Λ	Estimate (Std Err)
ω_{12}	-0.0629 (0.0288)	ξ_{12}	0.9749 (0.0064)	λ_{12}	0.1977 (0.0515)
ω_{13}	-0.0772 (0.0355)	ξ_{13}	0.9748 (0.0053)	λ_{13}	0.2230 (0.0472)
ω_{14}	-0.1388 (0.1928)	ξ_{14}	0.9878 (0.0109)	λ_{14}	0.2594 (0.2994)
ω_{15}	-0.0893 (0.0672)	ξ_{15}	0.9850 (0.0031)	λ_{15}	0.1976 (0.0973)
$\omega_{23 1}$	0.0191 (0.0071)	$\xi_{23 1}$	0.9521 (0.0145)	$\lambda_{23 1}$	0.0097 (0.0100)
$\omega_{24 1}$	0.0733 (0.0369)	$\xi_{24 1}$	0.8839 (0.0540)	$\lambda_{24 1}$	0.0311 (0.0161)
$\omega_{25 1}$	0.0332 (0.0117)	$\xi_{25 1}$	0.9375 (0.0162)	$\lambda_{25 1}$	0.0216 (0.0116)
$\omega_{34 12}$	0.0181 (0.0068)	$\xi_{34 12}$	0.9894 (0.0048)	$\lambda_{34 12}$	-0.0117 (0.0034)
$\omega_{35 12}$	0.0289 (0.0064)	$\xi_{35 12}$	0.9619 (0.0090)	$\lambda_{35 12}$	-0.0136 (0.0077)
$\omega_{45 123}$	0.0618 (0.0246)	$\xi_{45 123}$	0.9174 (0.0370)	$\lambda_{45 123}$	-0.0056 (0.0128)

Table 5: Diagonal QFDCC Model estimated by QML for portfolio I. The Bollerssev-Wooldridge standard deviations are in parentheses.

C^2	Estimate (Std Err)	A^2	Estimate (Std Err)	B^2	Estimate (Std Err)
c_{11}^2	0.0068 (0.0255)	a_{11}^2	0.0174 (0.0645)	b_{11}^2	0.9786 (0.0130)
c_{22}^2	0.0111 (0.0584)	a_{22}^2	0.0217 (0.1080)	b_{22}^2	0.9773 (0.0273)
c_{33}^2	0.0087 (0.0380)	a_{33}^2	0.0195 (0.2307)	b_{33}^2	0.9795 (0.0285)
c_{44}^2	0.0082 (0.0147)	a_{44}^2	0.0202 (0.0356)	b_{44}^2	0.9788 (0.0084)
c_{55}^2	0.0025 (0.0021)	a_{55}^2	0.0063 (0.0525)	b_{55}^2	0.9797 (0.0136)

Table 6: scalar DCC-GARCH Models estimated by QML for portfolio I. The Bollerssev-Wooldridge standard deviations are in parentheses.

Model	α	β
DCC	0.0284 (0.0032)	0.9674 (0.0041)

The same model is implemented for portfolio II, which is heterogenous in terms of geographical areas. Table 7 (resp. Table 9, Table 8) reports the estimation results of the C-Vine-GARCH (resp. scalar DCC, diagonal QFDCC) for portfolio II.

Table 7: Vine-GARCH Model estimated by QML for portfolio II: Germany (1), Greece (2), United-States (3), Italy (4), Japan (5), Australia (6). The Bollersev-Wooldridge standard deviations are in parentheses.

Ω	Estimate (Std Err)	Ξ	Estimate (Std Err)	Λ	Estimate (Std Err)
ω_{12}	0.0009 (0.0363)	ξ_{12}	0.9764 (0.0980)	λ_{12}	0.0473 (0.1015)
ω_{13}	0.0034 (0.0044)	ξ_{13}	0.9787 (0.0044)	λ_{13}	0.0421 (0.0080)
ω_{14}	-0.0637 (0.0258)	ξ_{14}	0.9795 (0.0043)	λ_{14}	0.1884 (0.0414)
ω_{15}	0.0059 (0.0041)	ξ_{15}	0.9714 (0.0127)	λ_{15}	0.0175 (0.0066)
ω_{16}	0.0045 (0.0036)	ξ_{16}	0.9772 (0.0047)	λ_{16}	0.0360 (0.0059)
$\omega_{23 1}$	-0.0064 (0.0225)	$\xi_{23 1}$	0.9388 (0.2172)	$\lambda_{23 1}$	0.0016 (0.0271)
$\omega_{24 1}$	0.0304 (0.1100)	$\xi_{24 1}$	0.8828 (0.4267)	$\lambda_{24 1}$	0.0092 (0.0350)
$\omega_{25 1}$	0.0080 (0.0074)	$\xi_{25 1}$	0.9601 (0.0211)	$\lambda_{25 1}$	0.0034 (0.0191)
$\omega_{26 1}$	0.0265 (0.0924)	$\xi_{26 1}$	0.9101 (0.2596)	$\lambda_{26 1}$	0.0121 (0.0497)
$\omega_{34 12}$	0.0015 (0.0035)	$\xi_{34 12}$	0.9551 (0.1663)	$\lambda_{34 12}$	0.0115 (0.0110)
$\omega_{35 12}$	-0.0001 (0.0003)	$\xi_{35 12}$	0.9942 (0.0055)	$\lambda_{35 12}$	0.0051 (0.0031)
$\omega_{36 12}$	-0.0008 (0.0016)	$\xi_{36 12}$	0.9805 (0.0356)	$\lambda_{36 12}$	0.0094 (0.0101)
$\omega_{45 123}$	0.0033 (0.0096)	$\xi_{45 123}$	0.7327 (0.2485)	$\lambda_{45 123}$	0.0128 (0.0217)
$\omega_{46 123}$	0.0035 (0.0031)	$\xi_{46 123}$	0.9512 (0.0191)	$\lambda_{46 123}$	0.0130 (0.0117)
$\omega_{56 1234}$	0.0134 (0.0067)	$\xi_{56 1234}$	0.9660 (0.0062)	$\lambda_{56 1234}$	0.0334 (0.0124)

Table 8: Diagonal QFDCC Model estimated by QML for portfolio II. The Bollersev-Wooldridge standard deviations are in parentheses.

C^2	Estimate (Std Err)	A^2	Estimate (Std Err)	B^2	Estimate (Std Err)
c_{11}^2	0.0065 (0.0029)	a_{11}^2	0.0139 (0.0061)	b_{11}^2	0.9851 (0.0025)
c_{22}^2	0.0012 (0.0016)	a_{22}^2	0.0021 (0.0026)	b_{22}^2	0.9931 (0.0026)
c_{33}^2	0.0020 (0.0036)	a_{33}^2	0.0029 (0.0054)	b_{33}^2	0.9876 (0.0029)
c_{44}^2	0.0064 (0.0050)	a_{44}^2	0.0134 (0.0103)	b_{44}^2	0.9856 (0.0028)
c_{55}^2	0.0021 (0.0091)	a_{55}^2	0.0021 (0.0097)	b_{55}^2	0.9925 (0.0041)
c_{66}^2	0.0067 (0.0172)	a_{66}^2	0.0086 (0.0231)	b_{66}^2	0.9904 (0.0030)

Table 9: scalar DCC/cDCC-GARCH Models estimated by QML for portfolio II. The Bollersev-Wooldridge standard deviations are in parentheses.

Model	α	β
DCC	0.0097 (0.0018)	0.9879 (0.0025)

For the sake of illustration, some series of correlations are reported in Section D in the appendix. Concerning C-vine GARCH models, the higher the level of the tree is, the smaller are the partial correlations coefficients ω and λ , for portfolio I. We may infer that once we control for the information given by Germany (1) (the core of the Eurozone) and United-Kingdom (2), estimating the dynamics of partial correlations on trees T_3 and T_4 is not necessary. This looks like evaluating white noise. This confirms the modeling of constrained vines, a case for which estimation results are close to the unconstrained case. On the contrary, this effect does not appear with the heterogenous portfolio II. Controlling for Germany, Greece and the US in portfolio II is not enough to deduce the whole information about the correlation dynamics between Japan and Australia, due to significant remaining idiosyncratic risks.

As for the DCC specification, the stationarity conditions are satisfied as $\hat{\alpha} + \hat{\beta} < 1$. All the estimated parameters are statistically significant at 1% level. The stationarity conditions also holds for the diagonal QFDCC by simply checking $a_{ii}^2 + b_{ii}^2 < 1$ for all $i = 1, \dots, n$. For both portfolios I and II, some parameters in C are not statistically significant. The estimated coefficients of A highlight how different both portfolios are. For portfolio I, the A coefficients are relatively close. But for portfolio II, they tend to be different, which justify the use of a diagonal QFDCC.

6.3 Specification testing

Once the model is estimated, we are able to forecast the variance-covariance matrix H_t , at least one-period ahead. There exist several methods to evaluate the absolute and/or relative efficiency of these predictions. See Patton and Sheppard (2007) for a survey. In this study, we focus on direct out-of-sample evaluation methods, which allow for pairwise comparisons. They test whether some or all the previous models provide different forecasts in terms of portfolio volatility behavior. Following the methodology of Engle and Colacito (2006) and Clements and al. (2009), we develop a mean-variance portfolio approach to test the variance covariance forecasts. The idea is the following: if a conditional variance covariance process is misspecified, then the minimum variance portfolio should emphasize the shortcoming. Then, we consider an investor who allocates at each time t his portfolio to the N assets according to a minimum-variance strategy. At each date t , he/she solves

$$\begin{cases} \min_{w_t} w_t' H_t w_t, \\ \text{s.t. } \iota' w_t = 1, \end{cases} \quad (14)$$

where w_t the $N \times 1$ vector of portfolio weights chosen at time $t - 1$, ι is a $N \times 1$ vector of 1 and H_t is the estimated conditional variance covariance matrix of the asset returns at time t , as deduced from some model dynamics that have been estimated on the sub-sample January 1999 - October 2011. Once the latter dynamic process is estimated, the predictions are plugged

into the program (14), for the sample November 2011 - August 2013. The solution of (14) is given by the global minimum variance portfolio:

$$w_t = \frac{H_t^{-1} \iota}{\iota' H_t^{-1} \iota}.$$

Engle and Colacito (2006) show that the realized portfolio volatility is the smallest one if the model covariance matrix is correctly specified. If one allocates his/her wealth using two different dynamic models i and j , whose predicted variance covariance matrices are (H_t^i) and (H_t^j) , the strategy providing the smallest portfolio variance will be considered as the best one. To do so, we consider a sequence of minimum variance portfolio weights $(w_{i,t})$ and $(w_{j,t})$ depending on the model. Then, we consider a distance based on the difference of the squared returns of the two portfolios:

$$u_{ij,t} = \{w'_{i,t} \epsilon_t\}^2 - \{w'_{j,t} \epsilon_t\}^2. \quad (15)$$

The portfolio variances are the same if the predicted variance covariance matrices are the same. Thus we test the null hypothesis

$$\mathcal{H}_0 : \mathbb{E}[u_{ij,t}] = 0$$

To do so, we use the test of Diebold and Mariano (1995), which consists of a least square regression using Heteroskedasticity Autoregressive Consistent standard errors:

$$\begin{aligned} u_{ij,t} &= \alpha + \epsilon_{u,t}, \quad \mathbb{E}[\epsilon_{u,t}] = 0, \\ \mathcal{H}_0 : \alpha &= 0. \end{aligned}$$

If the mean of $u_{ij,t}$ is positive (resp. negative), then the forecasts given by the covariance matrices of model j (resp. i) are better forecasts. We run the test for portfolios I and II for the scalar DCC, QFDCC, constrained C-Vine-GARCH (C-Vine-c) and unconstrained C-Vine-GARCH (C-Vine) models. The results are reported in tables 10 and 11. Those tables provide the out-of-sample t-statistics of the Diebold-Mariano test that checks the equality of a pair of series of covariance matrices using the loss function $u_{ij,t}$ defined by (15) over the period November 2011 - August 2013. This loss function is constructed as the difference of squared realized returns of alternative Multivariate GARCH models indicated in row i and column j . When the null hypothesis of equal predictive accuracy is rejected, a positive number is evidence in favor of the model in the column.

Table 10: Diebold Mariano Test of Multivariate GARCH models for portfolio I

	DCC	QFDCC	C-Vine	C-Vine-c
DCC		0.7217	0.0949	0.5864
QFDCC	-0.7217		-0.3157	0.2232
C-Vine	-0.0949	0.3157		2.2699**
C-Vine-c	-0.5864	-0.2232	-2.2699**	

Rejection of the nul hypothesis at: 10% for*, 5% for **, 1% for ***

Table 11: Diebold Mariano Test of Multivariate GARCH models for portfolio II

	DCC	QFDCC	C-Vine	C-Vine-c
DCC		-1.6014*	2.365e-4	-0.6087
QFDCC	1.6014*		1.6920*	0.9124
C-Vine	-2.365e-4	-1.6920*		-1.9671**
C-Vine-c	0.6087	0.9124	1.9671**	

Rejection of the null hypothesis at: 10% for*, 5% for **, 1% for ***

We first note that in the homogenous case, the DCC specifications do not provide better variance-covariance forecasts. Interestingly, the constrained case of the C-vine provides better prediction accuracy than the unconstrained case. For the heterogenous portfolio, we obtain the reverse. The C-vine specification outperforms the constrained case in terms of prediction accuracy: the two last levels of the tree should be estimated as once the dynamics are controlled by Germany, Greece and the US, there is still a significant idiosyncratic risk, which is not captured. Both versions of the C-vine are not outperformed by the scalar DCC, and the C-vine provides better variance-covariance forecasts than the QFDCC. The QFDCC is also slightly outperformed by the scalar DCC specification for the heterogenous portfolio, what is rather surprising. But all these results are not sufficiently clear-cut to draw any strong conclusion concerning a potential hierarchy between all these models. Other experiments should be surely necessary.

7 Conclusion

This paper has proposed to rely on vines to generate multivariate GARCH-type models. The main feature of our methodology is the specification/estimation of partial correlation processes “independently”, and their use to generate sequences of positive definite correlation matrices. The Canonical vine is particularly well-suited to model a hierarchy between asset returns, a bit as for factor models. Our approach does not rely on any normalization stage and we model directly correlation processes. Besides, the vine-GARCH approach allows for building parsimonious models, by setting constraints on the partial correlation dynamics, typically from some stage in the vine on. All these elements foster flexibility and enable to generate high-dimensional matrices.

The performances of the vine-GARCH and DCC estimators have been compared by means of applications to simulated and real data. The simulation study confirmed that a more flexible specification (the C-Vine-GARCH) provides a better accuracy. The constrained case is particularly adapted to homogenous portfolios and challenges the unconstrained case. The performances calculated from real data support the use of vine dynamics.

Therefore, a new framework has been opened in the field of MGARCH models. A challenging task for future research will be to provide the theoretical properties of this new family of estimators: stationarity, ergodicity, existence of moments, consistency and asymptotic normality. Due to the highly nonlinear feature of vine-GARCH models (through the F_{vine} function), this task is certainly very difficult. Beside theoretical open questions, more empirical work is probably necessary to evaluate all the advantages of such approaches w.r.t. more classical ones, as the standard DCC family.

References

- [1] Aas, K., Czado, C., Frigessi, A. and Bakken, H. (2006). Pair-copula constructions of multiple dependence. Working paper 487, Munich University.
- [2] Asai, M., McAleer, M. and Yu, J. (2006). Multivariate stochastic volatility: a review. *Econometric Reviews* 25, 145-175.
- [3] Bauwens, L., Laurent, S. and Rombouts, J. (2006). Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21, 79-109.
- [4] Bedford, T. and Cooke, R. (2002). Vines - a new graphical model for dependent random variables. *Ann. Statist.* 30(4), 1031-1068.
- [5] Bollerslev, T. and Wooldridge, J.M. (1994). Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time-Varying Covariances. *Econometric Reviews* 11, 143-172.
- [6] Bollerslev, T., Engle, R.F. and Wooldridge, J.M. (1988). A Capital Asset Pricing Model with Time-Varying Covariances. *Journal of Political Economy* 96, 116-131.
- [7] Bollerslev, T. (1990). Modeling the Coherence in Short-Run Nominal Exchange Rates: A Multivariate Generalized ARCH Model. *Review of Economics and Statistics* 72, 498-505.
- [8] Billio, M. and Caporin, M. (2006). A Generalized Dynamic Conditional Correlation Model for Portfolio Risk Evaluation. Working Paper Department of Economics, University of Venice.
- [9] Caporin, M. and McAleer, M. (2010). Ranking Multivariate GARCH Models by Problem Dimension. Social Science Research Network, Rochester, NY.
- [10] Caporin, M. and McAleer, M. (2013). Ten Things You Should Know about the Dynamic Conditional Correlation Representation. *Econometrics* 1, 115-126.
- [11] Campbell, J.Y. and Viceira, L.M. (2002). *Strategic Asset Allocation: Portfolio Choice for Long Term Investors*. Oxford University Press.
- [12] Campbell, J.Y. and Shiller, R.J. (1989) The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors. *The Review of Financial Studies* 1, 195-227.
- [13] Clements, A., Doolan, M., Hurn, S. and Becker, R. (2009). Evaluating multivariate volatility forecasts. NCER Working Paper Series No. 41. National Centre for Econometric Research.
- [14] Colacito, R., Engle, R.F. and Ghysels, E. (2011). A component model for dynamic correlations. *Journal of Econometrics* 164, 45-59.
- [15] Diebold, F.X. and Mariano, R.S. (1995). Comparing Predictive Accuracy. *Journal of Business and Economic Statistics* 13, 253-263.

- [16] Dissmann, J., Brechmann, E.C., Czado, C. and Kurowicka, D. (2012). Selecting and estimating regular vine copulae and application to financial returns. arXiv 1202.2002.
- [17] Engle R.F. (2002). Dynamic conditional correlation: a simple class of multivariate GARCH models. *Journal of Business and Economic Statistics* 20, 339-350.
- [18] Engle, R.F. and Sheppard, K. (2001). Theoretical and Empirical properties of Dynamic Conditional Correlation Multivariate GARCH. Working Paper No. 8554. National Bureau of Economic Research.
- [19] Fermanian, J.-D. and Malongo, H. (2013). Stationarity of Dynamic Conditional Correlation models. Working Paper Crest 2013-26.
- [20] Fermanian, J.-D. and Malongo, H. (2015). On the link between volatilities, regime switching probabilities and correlation dynamics. Working paper Crest.
- [21] Fiorentini, G., Calzolari, G., Panattoni, L. (1996). Analytic Derivatives and the Computation of Garch Estimates. *Journal of Applied Econometrics* 11, 399-417.
- [22] Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo Maximum Likelihood: Theory. *Econometrica* 52, 681-700.
- [23] Joe, H. (2006). Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis* 97, 2177-2189.
- [24] Kirkpatrick, S., Gelatt, C. D. and Vecchi, M. P. (1983). Optimization by Simulated Annealing. *Science, New Series* 220, 671-680.
- [25] Kurowicka, D. and Cooke, R. (2003). A parametrization of positive definite matrices in terms of partial correlation vines. *Linear Algebra and its Applications* 372, 225-251.
- [26] D. Kurowicka and R.M. Cooke (2006). Completion Problem With Partial Correlation vines. *Linear Algebra and its Applications* 418, 188-200.
- [27] Kurowicka, D. and Joe, H. (2010). Dependence modelling, vine copula handbook. World Scientific.
- [28] Kurowicka, D. (2011). Optimal truncation of vines. In *Dependence Modeling: Handbook on Vine Copulae*. World Scientific Publishing Co., 233-248.
- [29] Lewandowski, D., Kurowicka, D. and Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of Multivariate Analysis* 100, 1989-2001.
- [30] Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. and Teller, E. (1953). Equations of state calculations by fast computing machines. *The Journal of Chemical Physics* 21, 1087-1092.
- [31] Morales-Napoles, O., Cooke, R. and Kurowicka, D. (2009). The number of vines and regular vines on n nodes. *Discrete Applied Mathematics*.

- [32] Morales-Napoles, O. (2010). Bayesian belief nets and vines in aviation safety and other applications. PhD, University of Delft.
- [33] Nocedal, J., Wright, S.J. (2006). In *Numerical Optimization*. Springer Series in Operations Research.
- [34] Otranto, E. and Bauwens, L. (2013). Modeling the dependence of conditional correlations on volatility. WP Crenos.
- [35] Patton, A.J. and Sheppard, K. (2009). Evaluating Volatility and Correlation Forecasts. In Mikosch, T., Kreib, J.-P., Davis, R.A., Andersen, T.G. (Eds.) *Handbook of Financial Time Series*. Springer Berlin Heidelberg, 801-838.
- [36] Robert, C.P. and Casella, G. (1998). *Monte-Carlo Statistical Methods*. Springer.
- [37] Tse, Y.K. and Tsui, A.K.C. (2002). A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* 20, 351-362.
- [38] Veraverbeke, N., Omelka, M. and Gijbels, I. (2011). Estimation of a Conditional Copula and Association Measures. *Scandinavian Journal of Statistics* 38, 766-780.

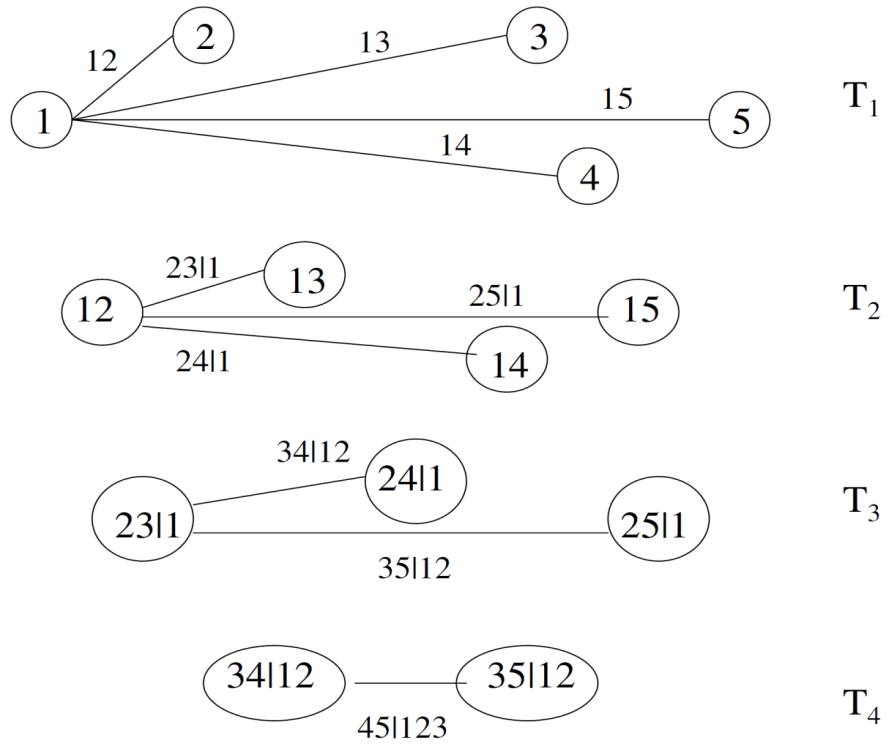


Figure 1: Example of a C-vine on five variables.

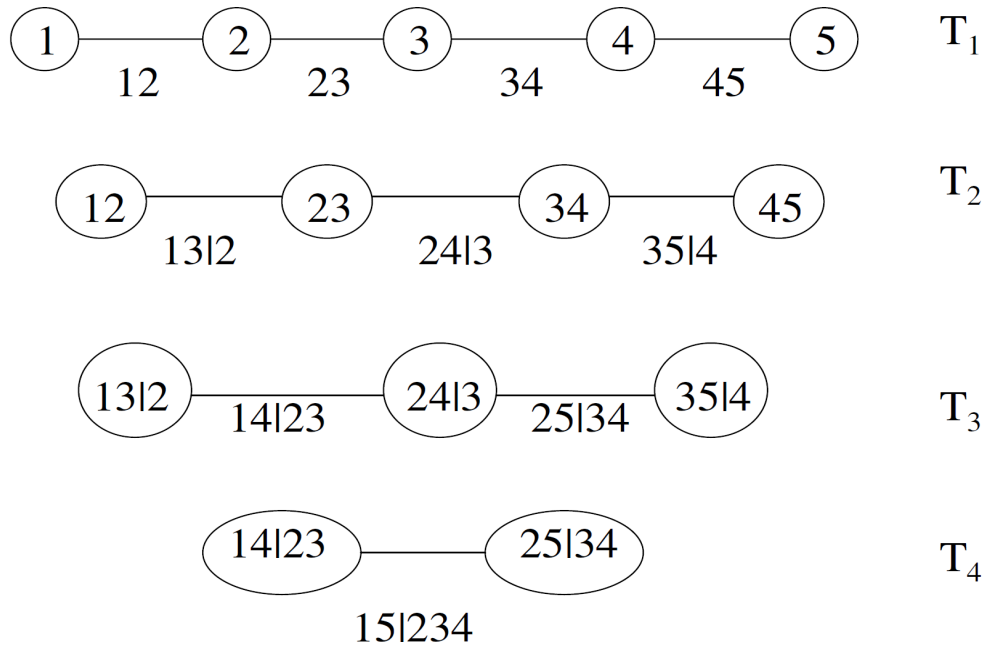


Figure 2: Example of a D-vine on five variables.

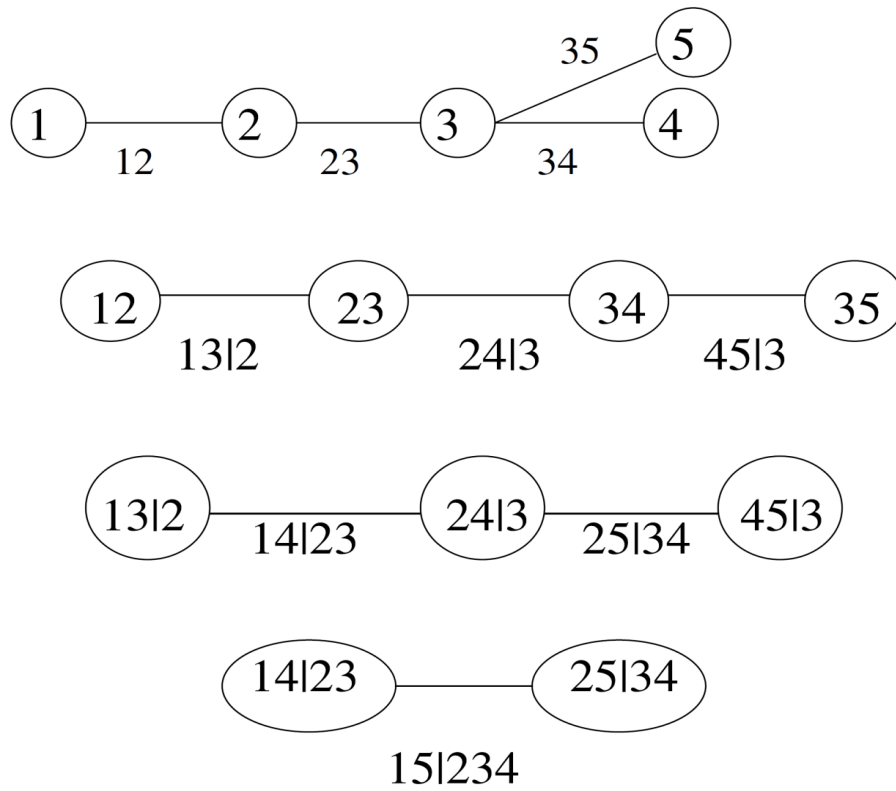


Figure 3: Example of a R-vine on five variables.

A Appendix: the detailed estimation formulae

In this appendix, we provide some details about the quasi-maximum estimation method. In particular, we suppose the distribution of the residuals is Gaussian and we make assumptions about the first and second order conditional moments. We split the likelihood function into two parts and apply a two step estimation procedure. To do so, the log likelihood function can be written as the sum of a volatility part and a correlation part

$$\ell(\theta; \epsilon) = -\frac{1}{2} \sum_{t=1}^T [N \log(2\pi) + \Delta(\epsilon_t; \theta_{vol}) + \Upsilon(\epsilon_t; \theta_{vol}, \theta_{cor}) + u_t' u_t],$$

with

$$\begin{aligned} \Delta(\epsilon_t; \theta_{vol}) &= \log(|D_t|) + \epsilon_t' D_t^{-1} \epsilon_t, \text{ and} \\ \Upsilon(\epsilon_t; \theta_{vol}, \theta_{cor}) &= \log(|R_t|) + u_t' R_t^{-1} u_t. \end{aligned}$$

A.1 First step estimation

Applying the quasi-maximum likelihood method, we get a non-linear system of equations defining the estimator of the volatility-related parameters, denoted by θ_{vol} . The volatility part of the log likelihood function is given by

$$QL1_T(\theta_{vol}) = -\frac{1}{2} \sum_{t=1}^T \sum_{i=1}^N \left[\log(h_{ii,t}) + \frac{\epsilon_{i,t}^2}{h_{ii,t}} \right].$$

$QL1_T$ corresponds to the sum of N univariate volatility processes. Thus we consider the method for one volatility process, say $QL1_{i,T}(\theta_{i,vol})$ with the parameter estimator $\theta_{i,vol}$ corresponding to the parameters of the volatility process of $r_{it} - \mu_{it}$. We model the conditional volatility process as a GARCH(1,1) process, such that

$$\begin{aligned} \forall i = 1, \dots, N, h_{ii,t} &= \varsigma_i + \kappa_i \epsilon_{i,t-1}^2 + \tau_i h_{ii,t-1}, \\ QL1_{i,T}(\theta_{i,vol}) &= -\frac{1}{2} \sum_{t=1}^T \left[\log(h_{ii,t}) + \frac{\epsilon_{i,t}^2}{h_{ii,t}} \right] := -\frac{1}{2} \sum_{t=1}^T l_{i,t}(\theta_{i,vol}). \end{aligned}$$

For convenience, we omit the i index of the parameter vector. The first order conditions for maximizing the log likelihood with respect to $\theta_{i,vol} = (\varsigma, \kappa, \tau)$ yield to the following orthogonality conditions:

$$\partial_{\theta_{i,vol}} QL1_{i,T}(\theta_{i,vol}) = -\frac{1}{2} \sum_{t=1}^T \left[\left(\frac{1}{h_{ii,t}} - \frac{\epsilon_{i,t}^2}{h_{ii,t}^2} \right) \partial_{\theta_{i,vol}} h_{ii,t} \right] = 0, \quad (16)$$

with $\partial_{\theta_{i,vol}} h_{ii,t} = \left(1, \epsilon_{i,t-1}^2, h_{ii,t-1}\right)' + \tau \partial_{\theta_{i,vol}} h_{ii,t-1}$. The system can be described as

$$\begin{cases} \partial_{\zeta} QL1_{i,T}(\theta_{i,vol}) &= -\frac{1}{2} \sum_{t=1}^T \left[\left(\frac{1}{h_{ii,t}} - \frac{\epsilon_{i,t}^2}{h_{ii,t}^2} \right) (1 + \tau \partial_{\zeta} h_{ii,t-1}) \right] = 0, \\ \partial_{\kappa} QL1_{i,T}(\theta_{i,vol}) &= -\frac{1}{2} \sum_{t=1}^T \left[\left(\frac{1}{h_{ii,t}} - \frac{\epsilon_{i,t}^2}{h_{ii,t}^2} \right) (\epsilon_{i,t-1}^2 + \tau \partial_{\kappa} h_{ii,t-1}) \right] = 0, \\ \partial_{\tau} QL1_{i,T}(\theta_{i,vol}) &= -\frac{1}{2} \sum_{t=1}^T \left[\left(\frac{1}{h_{ii,t}} - \frac{\epsilon_{i,t}^2}{h_{ii,t}^2} \right) (h_{ii,t-1} + \tau \partial_{\tau} h_{ii,t-1}) \right] = 0. \end{cases}$$

We rely on the Newton-Raphson method for solving this system. To do so, we apply a second order Taylor expansion of $QL1_{i,T}(\theta_{i,vol})$ around $\theta_{i,vol}^{(0)} \in \Theta$ as

$$\begin{aligned} QL1_{i,T}(\theta_{i,vol}) &\simeq QL1_{i,T}(\theta_{i,vol}^{(0)}) + \partial_{\theta_{i,vol}} QL1_{i,T}(\theta_{i,vol}^{(0)}) (\theta_{i,vol} - \theta_{i,vol}^{(0)}) + \\ &\quad \frac{1}{2} (\theta_{i,vol} - \theta_{i,vol}^{(0)})' \partial_{\theta_{i,vol} \theta_{i,vol}}^2 QL1_{i,T}(\theta_{i,vol}^{(0)}) (\theta_{i,vol} - \theta_{i,vol}^{(0)}). \end{aligned}$$

We minimize the quadratic function $QL1_{i,T}(\theta_{i,vol})$, which gives a vector $\theta_{i,vol}^{(1)}$ solution of the linear system

$$\partial_{\theta_{i,vol} \theta_{i,vol}}^2 QL1_{i,T}(\theta_{i,vol}^{(0)}) \theta_{i,vol}^{(1)} = \partial_{\theta_{i,vol} \theta_{i,vol}}^2 QL1_{i,T}(\theta_{i,vol}^{(0)}) \theta_{i,vol}^{(0)} - \partial_{\theta_{i,vol}} QL1_{i,T}(\theta_{i,vol}^{(0)}).$$

This is an iterative procedure. At iteration n , we build $QL1_{i,T}(\theta_{i,vol}^{(n)})$ as

$$\partial_{\theta_{i,vol} \theta_{i,vol}}^2 QL1_{i,T}(\theta_{i,vol}^{(n)}) (\theta_{i,vol}^{(n+1)} - \theta_{i,vol}^{(n)}) = -\partial_{\theta_{i,vol}} QL1_{i,T}(\theta_{i,vol}^{(n)}). \quad (17)$$

Hence the Newton-Raphson procedure requires the computation of the Hessian $\partial_{\theta_{i,vol} \theta_{i,vol}}^2 QL1_{i,T}(\theta_{i,vol})$, which is given by:

$$\begin{aligned} H &= \begin{pmatrix} \partial_{\zeta\zeta}^2 & \partial_{\zeta\kappa}^2 & \partial_{\zeta\tau}^2 \\ \partial_{\kappa\zeta}^2 & \partial_{\kappa\kappa}^2 & \partial_{\kappa\tau}^2 \\ \partial_{\tau\zeta}^2 & \partial_{\tau\kappa}^2 & \partial_{\tau\tau}^2 \end{pmatrix} QL1_{i,T}(\theta_{i,vol}) \\ &= -\frac{1}{2} \sum_{t=1}^T \left[\left(\frac{1}{h_{ii,t}} - \frac{\epsilon_{i,t}^2}{h_{ii,t}^2} \right) \partial_{\theta_{i,vol} \theta_{i,vol}}^2 h_{ii,t} + \left(\frac{2\epsilon_{i,t}^2}{h_{ii,t}^3} - \frac{1}{h_{ii,t}^2} \right) \partial_{\theta_{i,vol}} h_{ii,t} \partial_{\theta_{i,vol}}' h_{ii,t} \right]. \end{aligned}$$

Following Fiorentini and al. (1996), the Hessian is replaced by the empirical counterpart of the Fisher information matrix $\mathcal{I}(\theta_{i,vol}) = -\mathbb{E} \left[\partial_{\theta_{i,vol} \theta_{i,vol}}^2 l_{i,t}(\theta_{i,vol}) \right]$. This expression can be

simplified using conditional expectation:

$$\begin{aligned}
& -\mathbb{E}[\partial_{\theta_{i,vol}\theta'_{i,vol}}^2 l_{i,t}(\theta_{i,vol}) | \mathcal{F}_{t-1}] \\
&= \frac{1}{2} \mathbb{E} \left[\frac{1}{h_{ii,t}} \left(1 - \frac{\epsilon_{i,t}^2}{h_{ii,t}} \right) \partial_{\theta_{i,vol}\theta'_{i,vol}}^2 h_{ii,t} + \frac{1}{h_{ii,t}^2} \left(\frac{2\epsilon_{i,t}^2}{h_{ii,t}} - 1 \right) \partial_{\theta_{i,vol}} h_{ii,t} \partial_{\theta'_{i,vol}} h_{ii,t} | \mathcal{F}_{t-1} \right] \\
&= \frac{1}{2} \frac{1}{h_{ii,t}^2} \partial_{\theta_{i,vol}} h_{ii,t} \partial_{\theta'_{i,vol}} h_{ii,t}.
\end{aligned}$$

We used $\mathbb{E} \left[\frac{\epsilon_{i,t}^2}{h_{ii,t}} | \mathcal{F}_{t-1} \right] = 1$ and $h_{ii,t} \in \mathcal{F}_{t-1}$. Consequently, the Hessian is approximated by

$$\hat{\mathcal{I}}(\theta_{i,vol}) := -\frac{1}{2} \sum_{t=1}^T \frac{1}{h_{ii,t}^2} \mathbf{X}_{t-1,i} \mathbf{X}'_{t-1,i},$$

with $\mathbf{X}_{t,i} = \begin{pmatrix} 1 \\ \epsilon_{i,t}^2 \\ h_{ii,t} \end{pmatrix} + \tau \partial_{\theta_{i,vol}} h_{ii,t-1}$. The iterative numerical procedure is consequently given as

$$\begin{pmatrix} \varsigma^{(n+1)} \\ \kappa^{(n+1)} \\ \tau^{(n+1)} \end{pmatrix} = \begin{pmatrix} \varsigma^{(n)} \\ \kappa^{(n)} \\ \tau^{(n)} \end{pmatrix} + \left[\sum_{t=1}^T \frac{1}{h_{ii,t,(n)}^2} \mathbf{X}_{t-1,i,(n)} \mathbf{X}'_{t-1,i,(n)} \right]^{-1} \sum_{t=1}^T \frac{1}{h_{ii,t,(n)}} \mathbf{X}_{t-1,i,(n)} \left(\frac{\epsilon_{i,t}^2}{h_{ii,t,(n)}} - 1 \right), \quad (18)$$

with $\mathbf{X}_{t,i,(n)} = \begin{pmatrix} 1 \\ \epsilon_{i,t}^2 \\ h_{ii,t,(n)} \end{pmatrix} + \tau \partial_{\theta_{i,vol}} h_{ii,t-1,(n)}$. We initialize the latter algorithm by setting first

$\hat{h}_{ii,0} = \frac{1}{T-1} \sum_{t=1}^T \epsilon_{i,t}^2$, the empirical counterpart of the sample variance. Hence $\partial_{\theta_{i,vol}} \hat{h}_{ii,0} = 0$. As for the starting points of the sequence $(\theta_{i,vol,(n)})_n$, we set $\varsigma^{(0)} = 0$, $\kappa^{(0)} = 0.09$ and $\tau^{(0)} = 0.9$. The convergence criteria is such that $\|\tau^{(n+1)} - \tau^{(n)}\| \leq 10^{-8}$.

Finally, we estimate the variances of the estimators $\theta_{i,vol} = (\varsigma, \kappa, \tau)$ in a robust way using the formula of Bollerssev and Wooldridge (1992) as follows:

$$\begin{cases} \hat{\mathbb{V}}_{\hat{\varsigma}} &= \hat{V}_{\varsigma\varsigma} \\ \hat{\mathbb{V}}_{\hat{\kappa}} &= \hat{V}_{\kappa\kappa} \\ \hat{\mathbb{V}}_{\hat{\tau}} &= \hat{V}_{\tau\tau} \end{cases} \quad \text{with} \quad \begin{pmatrix} \hat{V}_{\varsigma\varsigma} & \hat{V}_{\varsigma\kappa} & \hat{V}_{\varsigma\tau} \\ \hat{V}_{\kappa\varsigma} & \hat{V}_{\kappa\kappa} & \hat{V}_{\kappa\tau} \\ \hat{V}_{\tau\varsigma} & \hat{V}_{\tau\kappa} & \hat{V}_{\tau\tau} \end{pmatrix} = A_T^{-1}(\hat{\theta}_{vol}) B_T(\hat{\theta}_{vol}) A_T^{-1}(\hat{\theta}_{vol}),$$

with

$$\begin{aligned}
A_T(\hat{\theta}_{i,vol}) &= \frac{1}{2T} \sum_{t=1}^T \partial_{\theta_{i,vol}} h_{ii,t-1} \partial_{\theta'_{i,vol}} h_{ii,t-1} \frac{1}{h_{ii,t}^2}, \\
B_T(\hat{\theta}_{i,vol}) &= \frac{1}{T} \sum_{t=1}^T \left(\frac{\epsilon_{i,t}^2}{2h_{ii,t}^2} - \frac{1}{2h_{ii,t}} \right)^2 \partial_{\theta_{i,vol}} h_{ii,t-1} \partial_{\theta'_{i,vol}} h_{ii,t-1}.
\end{aligned}$$

A.2 Second step estimation

Through the quasi-maximum likelihood method, a non-linear system of equations defines the an estimate of the parameter θ_{cor} , knowing θ_{vol} . Indeed, the correlation part of the log likelihood function is given by

$$QL2_T(\theta_{cor}; \hat{\theta}_{vol}, \epsilon) = -\frac{1}{2} \sum_{t=1}^T [\log(|R_t|) + u_t' R_t^{-1} u_t].$$

Due to the nullity of the score in expectation, we deduce the following orthogonality conditions:

$$\mathbb{E}[\partial_{\theta_{cor}} \Upsilon(\epsilon_t; \theta_{vol}, \theta_{cor})] = 0.$$

Taking the empirical counterpart of the previous quantity, we get

$$\sum_{t=1}^T \partial_{\theta_{cor}} [\log(|R_t|) + \hat{u}_t' R_t^{-1} \hat{u}_t] = 0,$$

where the individual volatility processes and the residuals \hat{u}_t have been obtained through the observations and the assumed parameter $\hat{\theta}_{cor}$. The analytic calculation of a score is tricky in the DCC-GARCH case, and unfeasible in the vine-GARCH case, which is highly non-linear. In the scalar DCC-GARCH specifications and the C-Vine iterative process, some algorithms using numerical scores and approximated Hessians can be used easily since we do not optimize criteria with numerous arguments. As for vine-GARCH models, when estimating without the iterative procedure and for reasonable portfolio sizes, we use a method that does not require any computation of a Gradient or Hessian: simulated annealing, as detailed in Subsection A.2.2.

A.2.1 The DCC-GARCH dynamics / The C-Vine iterative process

Similarly to the resolution of the first step, the previous system could be solved by the Newton-Raphson method using numerical derivatives. We apply the Sequential Quadratic Programming (SQP), which is widely examined by Nocedal and Wright (2006). We follow them and give in this section the main steps of the SQP approach. This method is used for DCC specifications and the C-Vine iterative process, a situation where we have to optimize in the a

subspace of \mathbb{R}^3 in practice. The SQP is a robust method in the case of constrained nonlinear problems. It is an iterative procedure, which considers a nonlinear problem as a quadratic programming subproblem, for a given value θ^n at iteration n and constructs a new iterate θ^{n+1} . In this setting, the sequence $(\theta^n)_n$ converges to a local minimum of the nonlinear problem as $n \rightarrow \infty$. The method is given here for the second step likelihood of the DCC specifications and one sub-likelihood in the C-Vine iterative process.

Our constrained nonlinear optimization problem is

$$\begin{cases} \arg \min_{\theta_{cor}} & \widetilde{QL2}_T(\theta_{cor}), \\ \text{subject to} & \phi(\theta_{cor}) \leq 0, \end{cases} \quad (19)$$

where $\phi(\cdot)$ is differentiable and $\widetilde{QL2}_T := -QL2_T$. Suppose the dimension of the parameter space for the correlation part is k . We associate a Lagrangian function to the nonlinear problem as

$$\begin{cases} \mathcal{L} : \mathbb{R}^k \times \mathbb{R}_+^m & \rightarrow \mathbb{R}, \\ \mathcal{L}(\theta_{cor}, \lambda) & = \widetilde{QL2}_T(\theta_{cor}) + \lambda\phi(\theta_{cor}), \end{cases} \quad (20)$$

where λ is the Lagrangian multiplier. The first order conditions of (20) at a local minimum correspond to the Karush-Kuhn-Tucker conditions. We assume that the second order sufficient optimality conditions given by Nocedal and Wright (2006) are satisfied. We consider quadratic programming subproblems, which should reflect the local properties of (19) with respect to the iterate θ_{cor}^n . To do so, we use a local quadratic approximation of the objective function and a local affine approximation of the constraint function:

$$\begin{aligned} \widetilde{QL2}_T(\theta_{cor}) &\approx \widetilde{QL2}_T(\theta_{cor}^n) + \partial_{\theta_{cor}} \widetilde{QL2}_T(\theta_{cor}^n)'(\theta_{cor} - \theta_{cor}^n) \\ &\quad + \frac{1}{2}(\theta_{cor} - \theta_{cor}^n)' \partial_{\theta_{cor}^2}^2 \widetilde{QL2}_T(\theta_{cor}^n)(\theta_{cor} - \theta_{cor}^n), \\ \phi(\theta_{cor}) &\approx \phi(\theta_{cor}^n) + \partial_{\theta_{cor}} \phi(\theta_{cor}^n)(\theta_{cor} - \theta_{cor}^n). \end{aligned}$$

We set $m(\theta_{cor}) = (\theta_{cor} - \theta_{cor}^n)$. The previous approximations lead to the subproblem

$$\begin{cases} \arg \min_{\theta_{cor}} & \partial_{\theta_{cor}} \widetilde{QL2}_T(\theta_{cor}^n)' m(\theta_{cor}) + \frac{1}{2} m(\theta_{cor})' \partial_{\theta_{cor}^2}^2 \widetilde{QL2}_T(\theta_{cor}^n) m(\theta_{cor}), \\ \text{subject to} & \phi(\theta_{cor}^n) + \partial_{\theta_{cor}} \phi(\theta_{cor}^n)' m(\theta_{cor}) \leq 0. \end{cases} \quad (21)$$

The numerical resolution of the latter system is led thanks to the Newton-Raphson method, which is given here with inequality constraints. To do so, we denote by λ^{n+1} the Lagrange multiplier for (21). We assume strict complementarity, a case where the multipliers are strictly positive at the optimum θ_{cor}^* . The idea of the method is the following one: if θ_{cor}^n and λ^n are sufficiently close to θ_{cor}^* and λ^* , then a solution of a quadratic problem with equality constraints also satisfies the sufficient conditions of a quadratic problem with inequality constraints. In

the latter case, the Karush-Kuhn-Tucker conditions are given by

$$\begin{cases} \partial_{\theta_{cor}} \widetilde{QL2}_T(\theta_{cor}^n) + \partial_{\theta_{cor}^2}^2 \widetilde{QL2}_T(\theta_{cor}^n) m(\theta_{cor}) + \lambda^{n+1} \partial_{\theta_{cor}} \phi(\theta_{cor}^n) & = 0, \\ \phi(\theta_{cor}^n) + \partial_{\theta_{cor}} \phi(\theta_{cor}^n)' m(\theta_{cor}) & = 0. \end{cases} \quad (22)$$

We define $m_\lambda = \lambda^{n+1} - \lambda^n$. Then (22) can be stated as

$$\begin{cases} -\partial_{\theta_{cor}} \mathcal{L}(\theta_{cor}^n, \lambda^n) & = \partial_{\theta_{cor}^2}^2 \widetilde{QL2}_T(\theta_{cor}^n) m(\theta_{cor}) + m_\lambda \partial_{\theta_{cor}} \phi(\theta_{cor}^n), \\ -\phi(\theta_{cor}^n) & = \partial_{\theta_{cor}} \phi(\theta_{cor}^n)' m(\theta_{cor}). \end{cases} \quad (23)$$

We minimize the Lagrangian function \mathcal{L} . Hence the Karush-Kuhn-Tucker conditions are given by $\Pi(\theta_{cor}, \lambda) = (\partial_{\theta_{cor}} \mathcal{L}(\theta_{cor}, \lambda), \phi(\theta_{cor}))' = 0$. To find the root of the latter equations, the Newton-Raphson method requires the Jacobian

$$J_{\theta_{cor}, \lambda} = \begin{pmatrix} \partial_{\theta_{cor}}^2 \mathcal{L}(\theta_{cor}, \lambda) & \partial_{\theta_{cor}} \phi(\theta_{cor}) \\ \partial_{\theta_{cor}} \phi(\theta_{cor})' & 0 \end{pmatrix}.$$

Hence the numerical resolution is given as

$$\begin{pmatrix} \partial_{\theta_{cor}}^2 \mathcal{L}(\theta_{cor}, \lambda) & \partial_{\theta_{cor}} \phi(\theta_{cor}) \\ \partial_{\theta_{cor}} \phi(\theta_{cor})' & 0 \end{pmatrix} \begin{pmatrix} \theta_{cor}^{n+1} - \theta_{cor}^n \\ \lambda^{n+1} - \lambda^n \end{pmatrix} = - \begin{pmatrix} \partial_{\theta_{cor}} \mathcal{L}(\theta_{cor}, \lambda) \\ \phi(\theta_{cor}) \end{pmatrix}. \quad (24)$$

We assume that the Hessian of \mathcal{L} is nonsingular and the Hessian of $\widetilde{QL2}_T$ is positive definite. Since $\partial_{\theta_{cor}^2}^2 \widetilde{QL2}_T(\theta_{cor}^n) = \partial_{\theta_{cor}^2}^2 \mathcal{L}(\theta_{cor}, \lambda)$, and using $\lambda^{n+1} = \lambda^n + m_\lambda$, we get $\partial_{\theta_{cor}} \phi(\theta_{cor}^n)' \lambda^{n+1} = \partial_{\theta_{cor}} \phi(\theta_{cor}^n)' (\lambda^n + m_\lambda)$. The solutions can be rewritten as

$$\begin{cases} \theta_{cor}^{n+1} & = \theta_{cor}^n - \left(\partial_{\theta_{cor}}^2 \widetilde{QL2}_T(\theta_{cor}^n) \right)^{-1} \partial_{\theta_{cor}} \mathcal{L}(\theta_{cor}^n, \lambda^{n+1}), \\ \lambda^{n+1} & = \lambda^n + \left(\partial_{\theta_{cor}} \phi(\theta_{cor}^n)' \partial_{\theta_{cor}}^2 \widetilde{QL2}_T(\theta_{cor}^n) \partial_{\theta_{cor}} \phi(\theta_{cor}^n) \right)^{-1} \left(\phi(\theta_{cor}^n) - \partial_{\theta_{cor}} \phi(\theta_{cor}^n)' \partial_{\theta_{cor}}^2 \widetilde{QL2}_T(\theta_{cor}^n) \partial_{\theta_{cor}} \phi(\theta_{cor}^n) \right). \end{cases} \quad (25)$$

For computing the Hessian, we apply a Broyden-Fletcher-Goldfarb-Shanno (BFGS) method, where the update at each iteration is given by:

$$\partial_{\theta_{cor}^2}^2 \widetilde{QL2}_T(\theta_{cor}^{n+1}) = \partial_{\theta_{cor}^2}^2 \widetilde{QL2}_T(\theta_{cor}^n) + \frac{x x'}{x' m(\theta_{cor}^{n+1})} - \frac{\partial_{\theta_{cor}^2}^2 \widetilde{QL2}_T(\theta_{cor}^n) m(\theta_{cor}^{n+1}) m(\theta_{cor}^{n+1})' \partial_{\theta_{cor}^2}^2 \widetilde{QL2}_T(\theta_{cor}^n)}{m(\theta_{cor}^{n+1})' \partial_{\theta_{cor}^2}^2 \widetilde{QL2}_T(\theta_{cor}^n) m(\theta_{cor}^{n+1})},$$

where $x = \partial_{\theta_{cor}} \mathcal{L}(\theta_{cor}^{n+1}, \lambda^{n+1}) - \partial_{\theta_{cor}} \mathcal{L}(\theta_{cor}^n, \lambda^{n+1})$. This updating procedure corresponds to a Quasi-Newton method. All the derivatives are computed thanks to finite differences.

The starting value θ_{cor}^0 of the correlation parameters are randomly drawn in a feasible set such that the inequality constraints are fulfilled. We choose $\partial_{\theta_{cor}^2}^2 \widetilde{QL2}_T(\theta_{cor}^0)$ as the identity matrix and, applying the first order condition to (19), the starting value for the lagrangian is

$$\lambda^0 = - \left(\partial_{\theta_{cor}} \phi(\theta_{cor}^0)' \partial_{\theta_{cor}} \phi(\theta_{cor}^0) \right)^{-1} \partial_{\theta_{cor}} \phi(\theta_{cor}^0)' \partial_{\theta_{cor}} \widetilde{QL2}_T(\theta_{cor}^0).$$

As for the inequality constraints contained in $\phi(\theta)$, we set the same bound constraints given by Engle and Sheppard (2001) for the DCC and Billio and Caporin (2006) for the diagonal QFDCC. For DCC type specifications, these constraints should ensure stationary and definite positive correlation matrix processes. When estimating the C-vine iterative process, we set $(\omega, \lambda) \in] - 0.999, 0.999]^2$ and $\xi \in]0, 0.999[$, which ensure non explosive partial correlation processes.

A.2.2 The vine-GARCH dynamics

Due to the high non-linearity of the vine-GARCH dynamics when estimated in one shot, we propose to apply a stochastic algorithm, the simulated annealing (SA), to maximize $QL2_T(\theta_{cor}; \hat{\theta}_{vol}, \epsilon) = -\frac{1}{2} \sum_{t=1}^T \Upsilon(\epsilon_t; \hat{\theta}_{vol}, \theta_{cor})$. To do so, we follow the method developed by Kirkpatrick and al. (1983), which is described by Robert and Casella (1998). This method consists of simulating the objective function from given values θ_{cor} . The key quantity is the Boltzman-Gibbs transform of $QL2_T(\theta_{cor}; \hat{\theta}_{vol}, \epsilon)$. It is

$$H_n(\theta_{cor}) \propto \exp \left\{ QL2_T(\theta_{cor}; \hat{\theta}_{vol}, \epsilon) / T_n \right\}, \quad (26)$$

where (T_n) is a positive decreasing sequence of so-called “temperatures”. As $T_n \rightarrow 0$, according to Theorem 5.2.7 in Robert and Casella (1998), the values that are simulated from (26) concentrate in a neighborhood closer and closer to a *local* maximum of $QL2_T$. But the SA is able to escape from local minima by applying a Metropolis simulation. Following Metropolis and al. (1953), we start from an initial point θ_{cor}^0 , and, for every $n = 0, 1, \dots$, we simulate a random variable from a *symmetric* uniform distribution $[-1, 1] / 100$. We denote its value by χ_n . Then the new value θ_{cor}^{n+1} is given by

$$\begin{cases} \theta_{cor}^{n+1} = \theta_{cor}^n + \chi_n & \text{with probability} \\ & p := \exp \left\{ \left(QL2_T(\theta_{cor}^n + \chi; \hat{\theta}_{vol}, \epsilon) - QL2_T(\theta_{cor}^n; \hat{\theta}_{vol}, \epsilon) \right) / T_n \right\} \wedge 1, \\ \theta_{cor}^{n+1} = \theta_{cor}^n & \text{with probability } 1 - p. \end{cases}$$

Therefore, if the move from χ_n increases $QL2_T$, i.e. $QL2_T(\theta_{cor}^n; \hat{\theta}_{vol}, \epsilon) \leq QL2_T(\theta_{cor}^n + \chi_n; \hat{\theta}_{vol}, \epsilon)$, then the new value is accepted. If $QL2_T(\theta_{cor}^n; \hat{\theta}_{vol}, \epsilon) \geq QL2_T(\theta_{cor}^n + \chi_n; \hat{\theta}_{vol}, \epsilon)$, the move can be kept with the acceptance probability p . Otherwise, a new random variable is simulated and then the acceptance test is run. This property enables the SA to escape from a local maximum with a probability which depends on the choice of the temperature T_n . The sequence (θ_{cor}^n) is an inhomogeneous Markov chain, because the probability depends on n . By theorem 5.2.4 of Robert and Casella (1998), there is still convergence of the chain. We normalize χ_n by 100 to explore a surface which is not too large, around the initial value of θ_{cor} . Hence we

allow for moves through $\theta_{cor}^n + \chi_n$, which are not too wide. As for the choice of the sequence T_n , we use a geometric rate, $T_n = \varkappa^n T_0$ with $\varkappa = 0.8$ and $T_0 = 5$.

A.2.3 Asymptotic variance covariance

We estimate the variances of the estimators $\mathbb{V}(\theta_{cor})$ in a robust way using the formula of Engle and Sheppard (2001). Due to the two step estimation procedure, this involves a trickier form than $A_T^{-1}(\hat{\theta}_{i,vol}) B_T(\hat{\theta}_{i,vol}) A_T^{-1}(\hat{\theta}_{i,vol})$, the asymptotic variance-covariance matrix of the univariate GARCH(1,1) processes. Recalling the asymptotic variance-covariance form of (12), the inverse of A_0 is

$$\begin{pmatrix} A_{11} & 0 \\ A_{12} & A_{22} \end{pmatrix}^{-1} = \begin{pmatrix} A_{11}^{-1} & 0 \\ A_{22}^{-1} A_{12} A_{11}^{-1} & A_{22}^{-1} \end{pmatrix}.$$

This yields the variance covariance of θ_{cor} :

$$\hat{\mathbb{V}}(\theta_{cor}) = -A_{22}^{-1} A_{12} A_{11}^{-1} B_{11} + A_{22}^{-1} B_{12} (-A_{22}^{-1} A_{12} A_{11}^{-1}) - A_{22}^{-1} A_{12} A_{11}^{-1} B_{12} A_{22}^{-1} + A_{22}^{-1} B_{22} A_{22}^{-1}. \quad (27)$$

The latter quantities can be estimated as follows:

$$\begin{aligned} A_{11} & \# \frac{1}{T} \sum_{t=1}^T \partial_{\theta_{vol} \theta_{vol}}^2 \Delta(\epsilon_t; \hat{\theta}_{vol}), \\ A_{12} & \# \frac{1}{T} \sum_{t=1}^T \partial_{\theta_{vol} \theta_{cor}}^2 \Upsilon(\epsilon_t; \hat{\theta}_{vol}, \hat{\theta}_{cor}), \\ A_{22} & \# \frac{1}{T} \sum_{t=1}^T \partial_{\theta_{cor} \theta_{cor}}^2 \Upsilon(\epsilon_t; \hat{\theta}_{vol}, \hat{\theta}_{cor}), \\ B_{11} & \# \frac{1}{T} \sum_{t=1}^T \partial_{\theta_{vol}} \Delta(\epsilon_t; \hat{\theta}_{vol})' \partial_{\theta_{vol}} \Delta(\epsilon_t; \hat{\theta}_{vol}), \\ B_{12} & \# \frac{1}{T} \sum_{t=1}^T \partial_{\theta_{vol}} \Delta(\epsilon_t; \hat{\theta}_{vol})' \partial_{\theta_{cor}} \Upsilon(\epsilon_t; \hat{\theta}_{vol}, \hat{\theta}_{cor}), \\ B_{22} & \# \frac{1}{T} \sum_{t=1}^T \partial_{\theta_{cor}} \Upsilon(\epsilon_t; \hat{\theta}_{vol}, \hat{\theta}_{cor})' \partial_{\theta_{cor}} \Upsilon(\epsilon_t; \hat{\theta}_{vol}, \hat{\theta}_{cor}). \end{aligned}$$

B Vine selection

In this section, we provide the sequential method to select an underlying C-vine. We rely on a sequential method that is a particular case of the general algorithm of Dissmann and al (2012). Note that the latter paper proposed a general algorithm to select the best R-vine

of a random vector. Here, the criteria will depend on the order of magnitude of empirical Kendall's tau and conditional Kendall's tau.

1. For tree T_1 and $N_1 = \{1, \dots, N\}$, maximize the dependence criterion

$$p_1 \leftarrow \arg \max_i \sum_{j \neq i} |\hat{\tau}_{ij}|,$$

where $\hat{\tau}_{ij}$ is the empirical Kendall's tau between assets i and j . Then, p_1 denotes the index of the variable which maximizes this criterion. The latter variable will be the root to build the edges on tree T_1 , which are the nodes on tree T_2 .

2. For $j = 2, \dots, N - 2$, $L_{j-1} = \{p_1, \dots, p_{j-1}\}$ and we maximize the dependence criterion:

$$p_j \leftarrow \arg \max_k \sum_{k, l \notin L_{j-1}, k \neq l} |\hat{\tau}_{kl|L_{j-1}}|,$$

where $\hat{\tau}_{kl|L_{j-1}}$ is the conditional Kendall's tau between k and l , knowing L_{j-1} (see below). Then, we set $L_j = L_{j-1} \cup \{p_j\}$.

At every stage, this sequential approach provides the variable which should enter into the conditioning set for the next tree.

In the program above, the empirical conditional and unconditional Kendall's tau have to be computed. As usual, the unconditional Kendall's tau is

$$\hat{\tau}_{ij} = \frac{4}{T(T-1)} \sum_{t=1}^T \sum_{t'=1}^T \mathbb{I}\{\epsilon_{i,t} < \epsilon_{i,t'}, \epsilon_{j,t} < \epsilon_{j,t'}\} - 1.$$

For the next trees, our approach requires the computation of conditional Kendall's tau. To do so, we use a nonparametric statistics proposed by Veraverbeke et al. (2011). Given a set of indices $L \subset \{1, \dots, N\}$, the conditional Kendall's tau of ϵ_i and ϵ_j given $\mathbf{x} = (x_k, k \in L)$ is

$$\hat{\tau}_{ij|L}(\mathbf{x}) = \frac{4}{1 - \sum_{t=1}^T w_{T,t}^2(\mathbf{x}, \mathbf{h})} \sum_{t=1}^T \sum_{t'=1}^T w_{T,t}(\mathbf{x}, \mathbf{h}) w_{T,t'}(\mathbf{x}, \mathbf{h}) \mathbb{I}\{\epsilon_{i,t} < \epsilon_{i,t'}, \epsilon_{j,t} < \epsilon_{j,t'}\} - 1,$$

where $w_{T,t}(\mathbf{x}, \mathbf{h}_T)$ denotes the Nadaraya-Watson weights given by

$$w_{T,t}(\mathbf{x}, \mathbf{h}) = \frac{K_{\mathbf{h}}(\vec{\epsilon}_t - \mathbf{x})}{\sum_{t'=1}^T K_{\mathbf{h}}(\vec{\epsilon}_{t'} - \mathbf{x})}, \quad t = 1, \dots, T,$$

with $\vec{\epsilon}_t = (\epsilon_{k,t}, k \in L)$ and $K_{\mathbf{h}}$ is a $\#L$ -variate kernel

$$K_{\mathbf{h}}(\vec{\epsilon}_t - \mathbf{x}) = \prod_{k \in L} \frac{1}{h_k} \phi\left(\frac{\epsilon_{k,t} - x_k}{h_k}\right),$$

with ϕ the density of a standardized Gaussian r.v. The vectorial bandwidth $\mathbf{h} \in \mathbb{R}_+^N$ is chosen with the usual rule-of-thumb $h_k = \hat{\sigma}_k/T^{1/5}$, where $\hat{\sigma}_k$ is the sample standard error of $(\epsilon_{k,t})_t$, $k \in L$. This conditional Kendall's tau is a function of \mathbf{x} . Therefore, we have to average over all possible values of the conditioning vector, to measure the degree of dependence of i and j knowing L . Here, this means we define $\hat{\tau}_{kl|L_{j-1}}$ above as

$$\hat{\tau}_{kl|L_{j-1}} := \frac{1}{M} \sum_{m=1}^M \hat{\tau}_{kl|L_{j-1}}(\mathbf{x}_m),$$

where \mathbf{x}_m is randomly chosen in $\{(\epsilon_{k,m}, k \in L_{j-1}), m = 1, \dots, T\}$ and the number of draws is $M = 500$.

C Appendix: estimation results of the constrained C-Vine dynamics for portfolio I and II

Table 12: C-Vine-GARCH Model estimated by QML for portfolio I: Germany (1), United-Kingdom (2), Italy (3), France (4), Netherlands (5). The Bollersev-Wooldridge standard deviations are in parentheses.

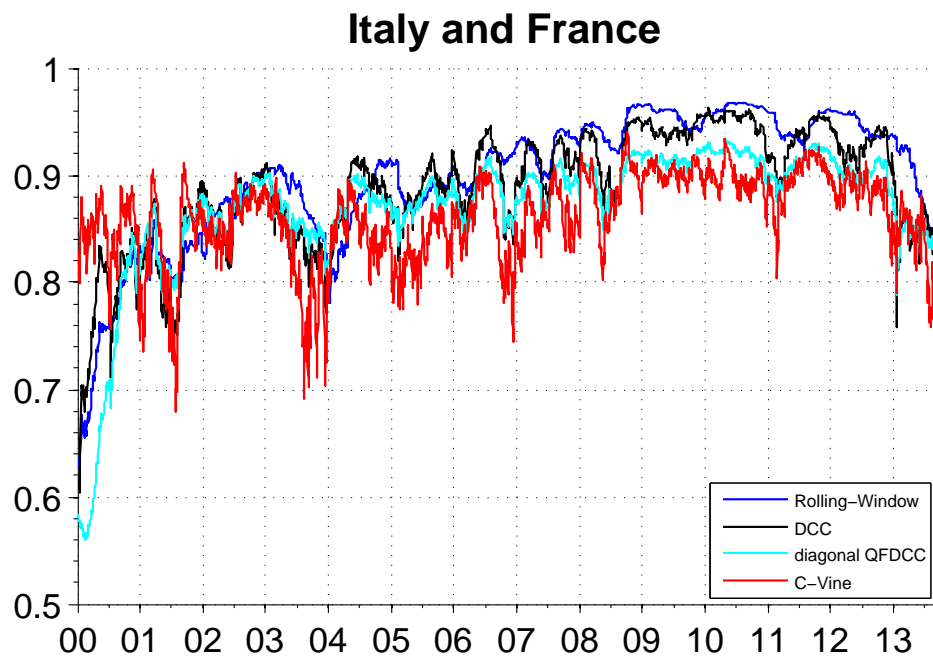
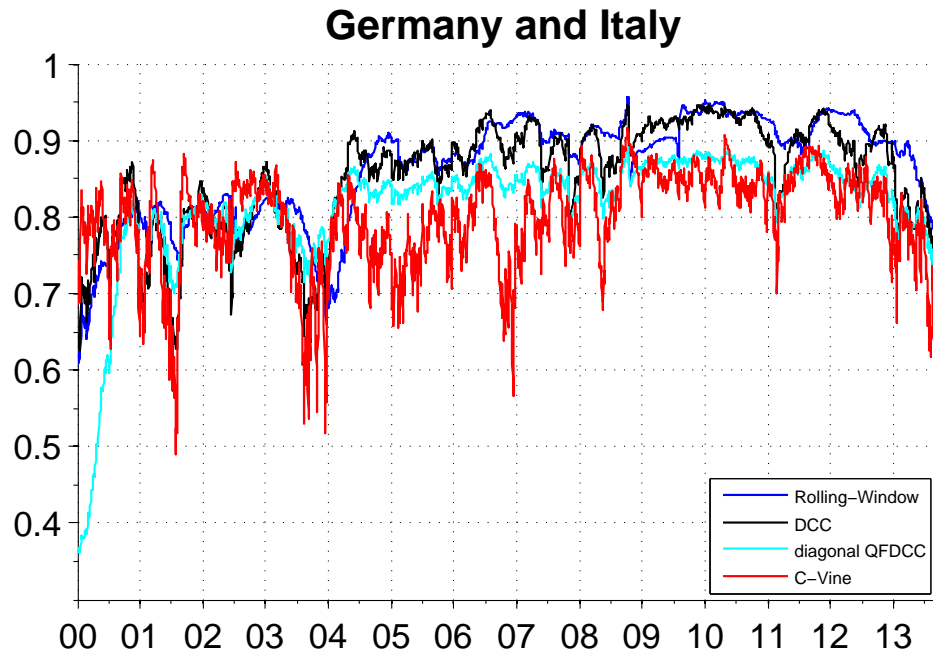
Ω	Estimate (Std Err)	Ξ	Estimate (Std Err)	Λ	Estimate (Std Err)
ω_{12}	-0.0453 (0.0182)	ξ_{12}	0.9802 (0.0062)	λ_{12}	0.1451 (0.0369)
ω_{13}	-0.0053 (0.0078)	ξ_{13}	0.9753 (0.0056)	λ_{13}	0.0791 (0.0176)
ω_{14}	-0.0319 (0.0166)	ξ_{14}	0.9862 (0.0081)	λ_{14}	0.0949 (0.0372)
ω_{15}	0077 (0.0047)	ξ_{15}	0.9715 (0.0143)	λ_{15}	0.0128 (0.0056)
$\omega_{23 1}$	0.0035 (0.0022)	$\xi_{23 1}$	0.9792 (0.0055)	$\lambda_{23 1}$	0.0086 (0.0105)
$\omega_{24 1}$	0.0154 (0.0087)	$\xi_{24 1}$	0.8399 (0.0844)	$\lambda_{24 1}$	0.0063 (0.0313)
$\omega_{25 1}$	0.0072 (0.0036)	$\xi_{25 1}$	0.9617 (0.0148)	$\lambda_{25 1}$	0.0090 (0.0114)
$\omega_{34 12}$	0.0066	$\xi_{34 12}$	-	$\lambda_{34 12}$	-
$\omega_{35 12}$	0.0583	$\xi_{35 12}$	-	$\lambda_{35 12}$	-
$\omega_{45 123}$	0.1166	$\xi_{45 123}$	-	$\lambda_{45 123}$	-

Table 13: Vine-GARCH Model estimated by QML for portfolio II: MSCI World (1), CAC 40 (2), German Bonds (3), Nikkei (4), S&P 500 (5), UBS-Gold (6). The Bollershev-Wooldridge standard deviations are in parentheses.

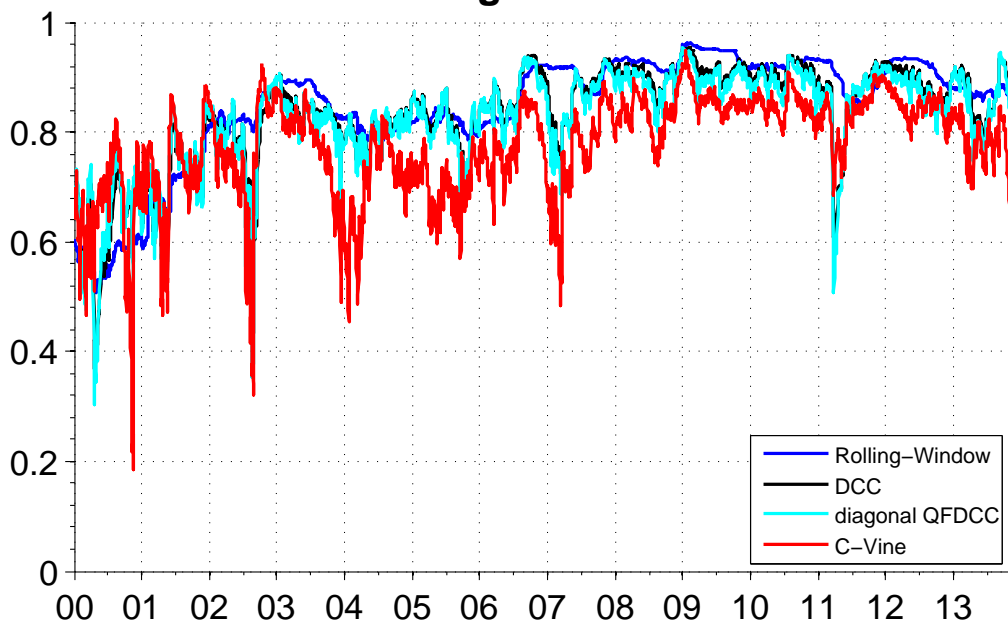
Ω	Estimate (Std Err)	Ξ	Estimate (Std Err)	Λ	Estimate (Std Err)
ω_{12}	0.0017 (0.0432)	ξ_{12}	0.9738 (0.1016)	λ_{12}	0.0513 (0.0974)
ω_{13}	0.0017 (0.0040)	ξ_{13}	0.9813 (0.0043)	λ_{13}	0.0398 (0.0081)
ω_{14}	-0.0610 (0.0254)	ξ_{14}	0.9796 (0.0044)	λ_{14}	0.1841 (0.0409)
ω_{15}	0.0082 (0.0047)	ξ_{15}	0.9679 (0.0124)	λ_{15}	0.0159 (0.0065)
ω_{16}	0.0055 (0.0046)	ξ_{16}	0.9743 (0.0059)	λ_{16}	0.0399 (0.0065)
$\omega_{23 1}$	-0.0081 (0.0268)	$\xi_{23 1}$	0.9203 (0.2694)	$\lambda_{23 1}$	0.0008 (0.0349)
$\omega_{24 1}$	0.0324 (0.0992)	$\xi_{24 1}$	0.8722 (0.3983)	$\lambda_{24 1}$	0.0094 (0.0400)
$\omega_{25 1}$	0.0091 (0.0098)	$\xi_{25 1}$	0.9555 (0.0303)	$\lambda_{25 1}$	0.0050 (0.0160)
$\omega_{26 1}$	0.0089 (0.0318)	$\xi_{26 1}$	0.9675 (0.0832)	$\lambda_{26 1}$	0.0089 (0.0303)
$\omega_{34 12}$	0.0013 (0.0036)	$\xi_{34 12}$	0.9585 (0.1716)	$\lambda_{34 12}$	0.0109 (0.0095)
$\omega_{35 12}$	-0.0001 (0.0009)	$\xi_{35 12}$	0.9936 (0.0156)	$\lambda_{35 12}$	0.0060 (0.0090)
$\omega_{36 12}$	-0.0004 (0.0030)	$\xi_{36 12}$	0.9907 (0.0911)	$\lambda_{36 12}$	0.0057 (0.0279)
$\omega_{45 123}$	0.0311	$\xi_{45 123}$	-	$\lambda_{45 123}$	-
$\omega_{46 123}$	0.2424	$\xi_{46 123}$	-	$\lambda_{46 123}$	-
$\omega_{56 1234}$	0.7142	$\xi_{56 1234}$	-	$\lambda_{56 1234}$	-

D Appendix: Conditional correlation trajectories

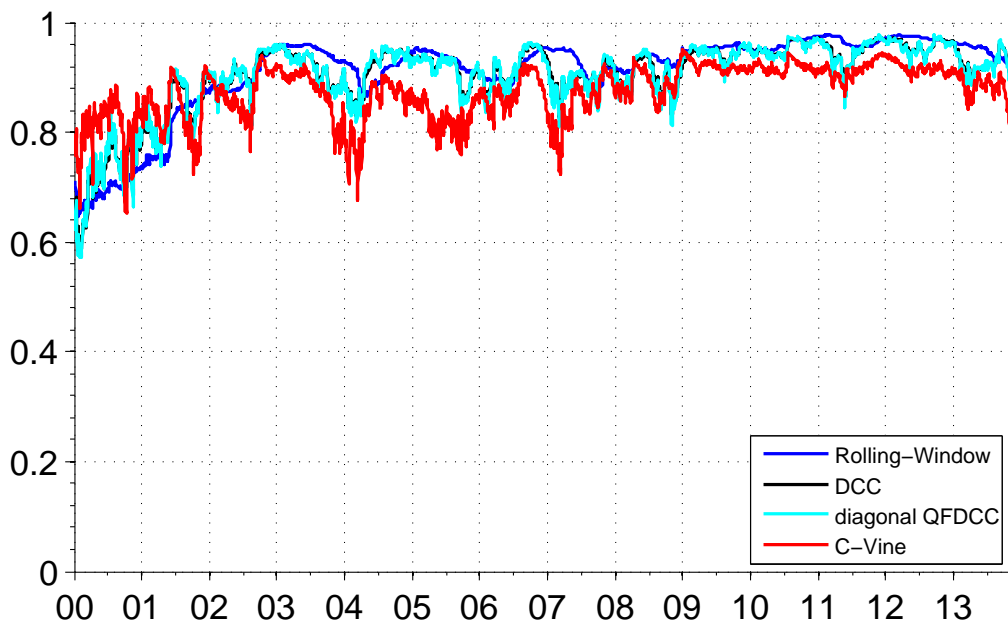
D.1 Portfolio I



United-Kingdom and France

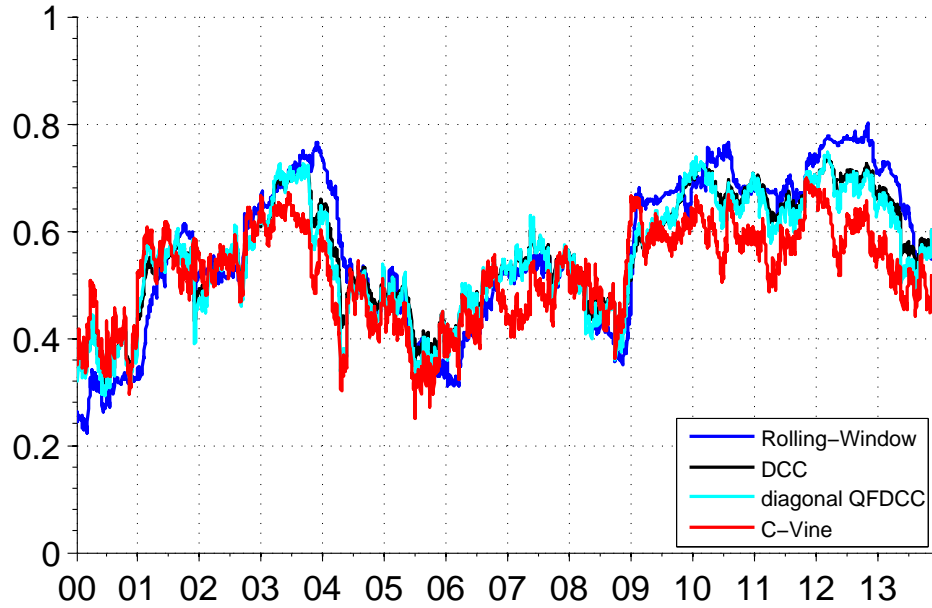


France and Netherlands

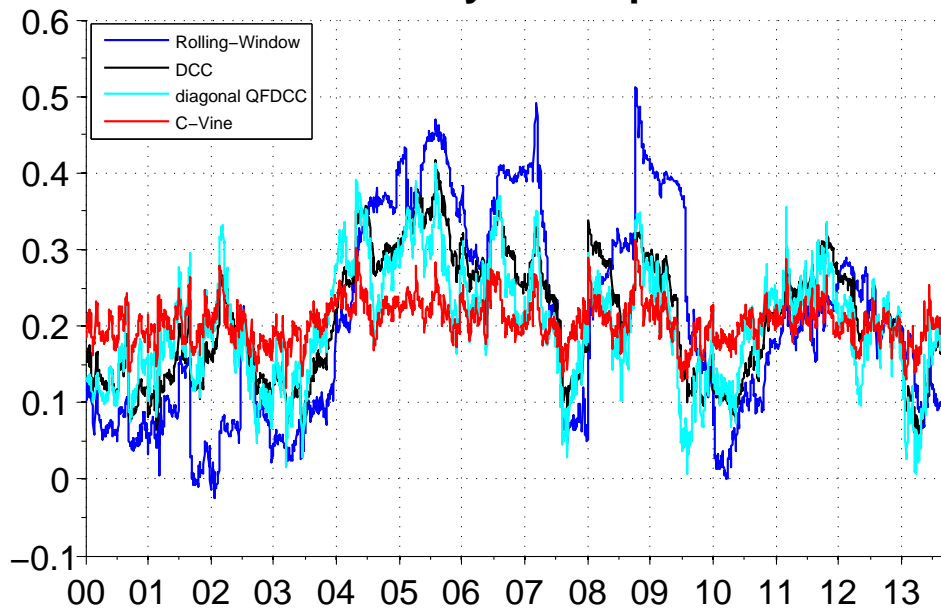


D.2 Portfolio II

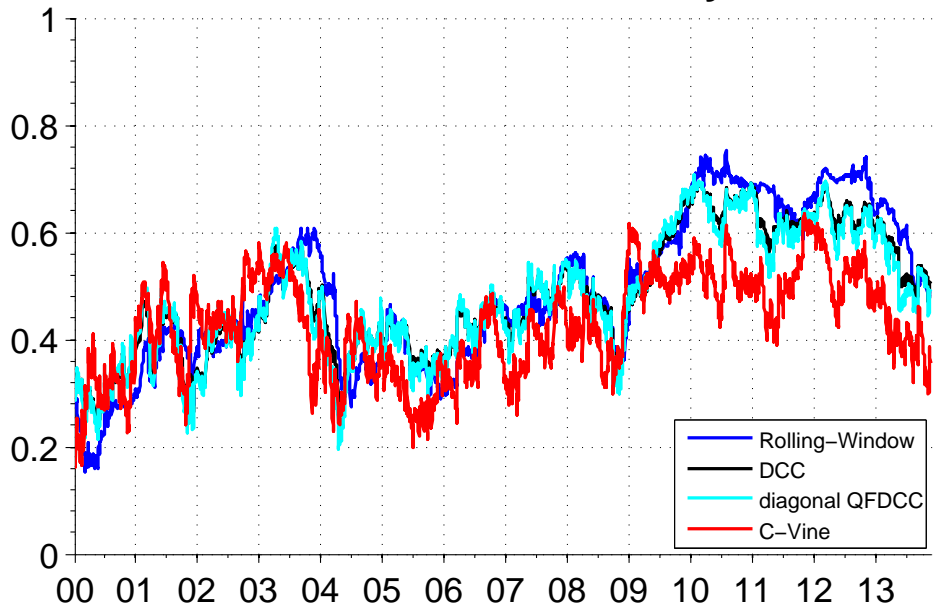
Germany and United-States



Germany and Japan



United-States and Italy



Italy and Japan

