Econometric Theory, 2018, Page 1 of 31.
doi:10.1017/S026646661800004X

DYNAMIC ASSET CORRELATIONS BASED ON VINES

BENJAMIN POIGNARD
Osaka University

JEAN-DAVID FERMANIAN
Crest

We develop a new method for generating dynamics of conditional correlation matrices of asset returns. These correlation matrices are parameterized by a subset of their partial correlations, whose structure is described by a set of connected trees called "vine". Partial correlation processes can be specified separately and arbitrarily, providing a new family of very flexible multivariate GARCH processes, called "vine-GARCH" processes. We estimate such models by quasi-maximum likelihood. We compare our models with DCC and GAS-type specifications through simulated experiments and we evaluate their empirical performances.

1. INTRODUCTION

A multivariate setting is necessary for modeling the cross-sectional and temporal dependencies between N financial asset returns. The usual modeling approach relies on the specification of the first two moments of vectors of returns conditional on their past (and current market information possibly). Once this is done, some assumed vectors of innovations close the model specification. The multivariate GARCH (MGARCH) and the multivariate stochastic volatility (MSV) models are the two main frameworks: see the surveys of Bauwens, Laurent, and Rombouts (2006) and Asai, McAleer, and Yu (2006), respectively. Such approaches allow for generating sequences of asset return covariance matrices (H_t), and then provide their correlations as a by-product. In financial econometrics, MGARCH models are most commonly used: Factor-GARCH (Engle, Ng, and Rothschild, 1990), BEKK (Engle and Kroner, 1995), DCC (Engle, 2002) models, among numerous competitors. They induce some typical patterns such as volatility clustering and

The authors are grateful for very helpful discussions with Christian Francq and Jean-Michel Zakoïan. They thank numerous seminar participants, particularly at the 11th World Congress of the Econometric Society, Computational and Financial Econometrics 2015, the Bachelier seminar, the Oberwolfach copulae conference 2015, etc. Numerous remarks of the AE and of two anonymous referees have allowed significant improvements of the article. The authors have been supported by the labex Ecodec. Address correspondence to Benjamin Poignard, Graduate School of Engineering Science, Osaka University, 1-3 Machikaneyama, Toyonaka, Osaka 560-0043, Japan; e-mail: poignard@sigmath.es.osaka-u.ac.jp and Jean-David Fermanian, Crest, 5 Avenue Henry le Chatelier, TSA 26644, 91764 Palaiseau Cedex, France; e-mail: jean-david.fermanian@ensae.fr.

complex dependencies. They are estimated through QML, without the necessity of complex inference procedures, contrary to most MSV models.

Nowadays, MGARCH models are faced with three main challenges:

(i) to guarantee the positive definiteness of the generated conditional covariance matrices;
(ii) to propose flexible and realistic model specifications;
(iii) when the number of underlying assets is "large" (more than five, in practice), to be able to estimate such models in practice.

Since the number of MGARCH parameters dramatically increases with the number of underlying assets, some simplified MGARCH specifications have searched for parsimony, but sometimes at the price of an over-simplification: the scalar DCC process of Engle and Sheppard (2001), the Flexible DCC model of Billio and Caporin (2006), among others. Unfortunately, the ability of the latter models to capture complex and rich dynamics of heterogeneous series is limited. Besides, the necessity of obtaining (positive definite) covariance matrices has induced some more or less ad-hoc normalization stages (DCC) or algebraic/nonlinear transforms (Cholevski-, Orthogonal-GARCH). Such matrix transforms make the interpretation of parameter values harder, to finely explain correlation dynamics. Indeed, neither standard MGARCH nor DCC-type models work directly on explicit correlation dynamics.

Our aim is to remain inside the MGARCH family, without suffering from these drawbacks. In this article, we propose to circumvent the problems (i)–(iii) above,

- by managing time-varying volatilities and correlations separately, as in DCC-GARCH models,
- by directly specifying partial correlation dynamics.

This approach tends to be both parsimonious and flexible. 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 partial correlations (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 are. This property will be crucial: by producing univariate dynamics of partial correlations independently, we automatically obtain sequences of correlation matrices, solving (i).

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 mapped to a so-called vine tree. Such objects are sets of connected trees (see Appendix 2). They have been discovered recently due to their ability to build high-dimensional distributions through a set of bivariate copulas (one copula per node of the vine) and marginal cdfs'. See Aas,

Czado, Frigessi, and Bakken (2006) for an introduction. Here, we develop a class of MGARCH models based on regular vines, the so-called "vine-GARCH" models. These models are flexible enough by allowing independent specifications of partial correlation processes. They also foster parsimony as one can set constraints at any level of the vine tree, answering point (ii). Estimation can be led equation-by-equation through an iterative procedure on the nodes of the vine, solving (iii). Note that Creal, Koopman, and Lucas (2011) have proposed an alternative parameterization of correlation matrices, based on dynamic angles defining hyperspherical coordinates. These parameters are free of constraints, as in our case. However, these coefficients do not enjoy a direct and intuitive interpretation contrary to partial correlations, despite the attempts of Zhang, Leng, and Tang (2015).

This article is organized as follows[1]: the new vine-GARCH framework is detailed in Section 2. In Section 3, we define the statistical inference of our new models by a quasi maximum likelihood (QML) procedure. Section 4 contains an empirical study with simulated data and a database of stock returns, and then we conclude the study.

2. VINE-GARCH CORRELATION DYNAMICS

2.1. The Usual Multivariate-GARCH Framework

We denote by $(\epsilon_t)_{t=1,\dots,T}$ a sequence of N-dimensional random vectors, whose dynamics is specified by θ, a finite-dimensional parameter. Denote by $(\mathcal{F}_t)$ the natural filtration, i.e., $\mathcal{F}_t := \sigma(\epsilon_s, s \leq t)$ and $\mathbb{E}_{t-1}[X] := \mathbb{E}[X|\mathcal{F}_{t-1}]$ for any random quantity X. The key model assumption is

$$\epsilon_t = H_t^{1/2}(\theta)\,\eta_t, \tag{1}$$

where the series $(\eta_t)_{t\geq 1}$ is supposed to be a strong white noise s.t. $\mathbb{E}[\eta_t] = 0$ and $\mathrm{Var}(\eta_t) = I_N$. We suppose $H_t(\theta) := H_t := \mathrm{Var}_{t-1}(\epsilon_t)$ is a $N \times N$ positive definite matrix. At this stage, the model is semiparametric. Its specification is complete when the law of η_t and the dynamics of $(H_t(\theta))$ are specified. In this article, we mainly focus on the latter point.

The matrix H_t represents the unobserved time-dependent conditional covariance matrix of the process (ϵ_t). In most models, a brute-force inference seems to be infeasible even when the dimension N is small. To avoid this problem, a common approach consists of splitting the problem into two simpler ones: modelling conditional volatilities on one side, the correlation dynamics on the other side, as for DCC models.

Denote by $h_{i,t}$ the conditional variances of $(\epsilon_{i,t})$ and $\rho_{ij,t}$ the conditional correlations between $\epsilon_{i,t}$ and $\epsilon_{j,t}$, for $i, j = 1, \dots, N$, $i < j$. In matrix notation, $H_t = D_t R_t D_t$ where $D_t = \mathrm{diag}\left(h_{1,t}^{1/2}, \dots, h_{N,t}^{1/2}\right)$ is the diagonal matrix of the conditional volatilities, and $R_t = \left[\rho_{ij,t}\right]$ is the matrix of the conditional correlations. By construction, R_t is the conditional covariance matrix of the vector of the standardized returns $u_t = \left(u_{1,t}, \dots, u_{N,t}\right)$ with $u_{i,t} = \epsilon_{i,t}/\sqrt{h_{i,t}}$. Both volatility

and correlation dynamics depend on a specific set of parameters given by $\theta = (\theta_v, \theta_c)' \in \Theta_v \times \Theta_c$, where θ_v (resp. θ_c) is the set of parameters determining the volatility processes (resp. correlation process).

Let us assume that, for every $i = 1, \ldots, N$ and t, there exists a function h_i s.t.

$$h_{i,t} = h_i \left(\theta_v^{(i)}; \epsilon_{i,t-1}, \ldots, \epsilon_{i,t-q_i}; h_{i,t-1}, \ldots, h_{i,t-p_i} \right), \tag{2}$$

for some positive integers p_i and q_i and some parameter $\theta_v^{(i)} \in \mathbb{R}^{p_i+q_i+1}$. Once stacked, the parameters $\theta_v^{(i)}$ provide θ_v. Typically, we could assume GARCH(p_i,q_i) processes in (2), or even other univariate GARCH-type models (EGARCH, GJR-GARCH, T-GARCH, etc). Since our vine-GARCH framework only needs consistent estimates of conditional volatilities, as deduced from this first stage, there is a high degree of flexibility to specify the individual volatility dynamics.

Note that we have supposed no spill-over effects between different asset volatilities in equation (2). This assumption simplifies the estimation of θ_v by allowing an equation-by-equation inference procedure, and it is almost unavoidable when N is large. This absence of spill-over effects is commonly used in the DCC literature, even if it may be questionable. Indeed, some studies have exhibited significant spill-over effects empirically: see Hamao, Masulis, and Ng (1990), Koutmos and Booth (1995), Liao and Williams (2004), among others. We stress that this point is not crucial for our vine-GARCH model, and this assumption could be removed: see Remark 2 below.

2.2. Vines and Partial Correlations

Let $\mathbf{X} = (X_1, \ldots, X_n)$ be a n-dimensional random vector, $n \geq 2$, with zero mean. For any indices i, j in $\{1, \ldots, n\}$, $i \neq j$ and any subset $L \subset \{1, \ldots, n\}$, for which i and j do not belong to L, $\rho_{i,j|L}$ is called the partial correlation of X_i and X_j, given X_k, $k \in L$. It is the correlation between the orthogonal projections of X_i and X_j on $< X_k, k \in L >^{\perp}$, the orthogonal of the subspace generated by $\{X_k, k \in L\}$. When L is empty, then $\rho_{i,j|\emptyset} = \rho\left(X_i, X_j\right) := \rho_{i,j}$ is the usual correlation. Note that, if the random vector $\mathbf{X}$ is normal, then its partial correlations correspond to some conditional correlations.

Interestingly, partial correlations can be computed from usual correlations with a recursive formula. Let (i, j, k) be any set of distinct indices, and L be another (possibly empty) set of indices that is disjoint from (i, j, k). Following Lewandowski et al. (2009), we have

$$\rho_{i,j|k,L} = \frac{\rho_{i,j|L} - \rho_{i,k|L}\rho_{j,k|L}}{\sqrt{\left(1 - \rho_{i,k|L}^2\right)\left(1 - \rho_{j,k|L}^2\right)}}. \tag{3}$$

Assume we know the usual correlations $\rho_{i,j}$, for any couple (i, j), $i \neq j$. We check easily that any partial correlation can be calculated by invoking (3) several times with increasing subsets L. Actually, the opposite property is true if

we start from a convenient subset of partial correlations, given by a regular vine or R-vine (see the appendix for a basic toolkit concerning the necessary vine concepts). Indeed, the edges of a regular vine on n elements may be associated with the partial correlations of a n-dimensional random vector in the following way: for $i = 1, \ldots, n-1$, consider any edge e in the vine. 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 edge. Kurowicka and Cooke (2006) call this structure a *partial correlation vine specification*, that is simply an R-vine for which any edge is associated with a number in $]-1,1[$. To illustrate these ideas, consider the vine of Figure 1 under a partial correlation point of view: an associated partial correlation vine will specify the set of partial correlations $\{\rho_{12}, \rho_{13}, \rho_{14}, \rho_{15}, \rho_{23|1}, \rho_{24|1}, \rho_{25|1}, \rho_{34|12}, \rho_{35|12}, \rho_{45|123}\}$, that is sufficient to recover the correlation matrix between the five assets.

Actually, all positive definite correlation matrices may be generated by setting an R-vine (fixed) on n variables, and by assigning different partial correlations to all the edges of this vine. This means setting ρ_e to any edge e, and these partial correlations may be chosen in $]-1,1[$ *arbitrarily*. This is the content of Corollary 7.5 in Bedford and Cooke (2002).

THEOREM 1 (Bedford and 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 for the vine.

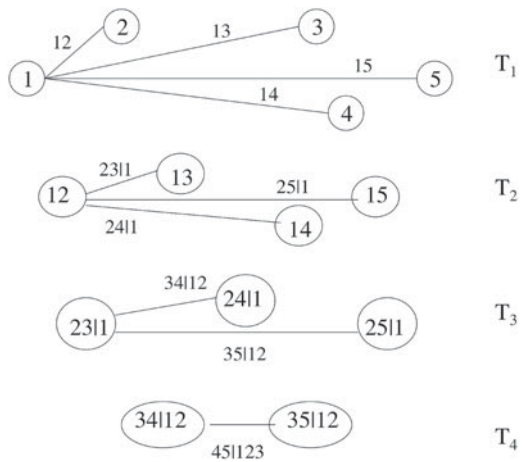

FIGURE 1. Example of a C-vine on five variables. The two nodes $(1,2)$ and $(1,3)$ in T_2 are connected by the edge $(2,3|1)$, whose constraint set is $\{1,2,3\}$, conditioned set is $\{2,3\}$ and conditioning set is $\{1\}$.

In other words, we automatically generate a positive definite correlation matrix out of *any set* of partial correlations provided by *any R-vine*. Conversely, for any fixed R-vine, every correlation matrix can be recovered from a convenient set of partial correlations that is associated with the edges of this vine. Actually, the formulas (3) above enable to build such $n \times n$ correlation matrices based on $n(n-1)/2$ *arbitrarily chosen partial correlations* (see Kurowicka and Cooke, 2003, or Joe, 2006). For a given partial correlation vine, some explicit algorithms map the (usual) correlations and the underlying partial correlations: see Lewandowski et al. (2009). Such algorithms are available in the R-package called "vine-copula" (see Brechmann and Schepsmeier, 2013, for instance). Daniels and Pourahmadi (2009) use a similar result to the formulas (3) to parameterize correlation matrices with partial autocorrelations. Nonetheless, they consider static random vectors (no multivariate dynamics) and under a Bayesian point of view.

2.3. Our Model Specification

Given the dynamics of the conditional variances, we now focus on the dynamics of the conditional correlations, i.e., the process (R_t). The latter one will be deduced from dynamic models of partial correlations. We use a partial correlation vine specification, i.e., a given regular vine $V(N)$ and $N(N-1)/2$ numbers in $]-1,1[$ to specify the corresponding partial correlations. And we invoke 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 Pc_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 "partial correlation" dynamics:

$$H_t = D_t R_t D_t, \tag{4}$$

$$\Psi(Pc_t) = \Omega + \sum_{k=1}^{p} \Xi_k \Psi(Pc_{t-k}) + \sum_{l=1}^{q} \Lambda_l \zeta_{t-l}, \tag{5}$$

$$R_t = \text{vechof}\left(F_{\text{vine}}(Pc_t)\right), \quad \text{where} \tag{6}$$

- The vector Pc_t is the "partial correlation vector" deduced from the given R-vine structure. We denote $Pc_t := (\rho_{ij|L,t}, (ij|L) \text{ are edges of } V(N))'$, when the underlying vine is $V(N)$.
- The Ξ_k and Λ_l denote $N(N-1)/2 \times N(N-1)/2$ matrices of unknown parameters, and Ω is a $N(N-1)/2$ unknown vector. Set the vector of parameters $\theta_c = \left(\Omega, \Xi_1, \ldots, \Xi_p, \Lambda_1, \ldots, \Lambda_q\right)$. Note that these matrices are *arbitrarily chosen*, and we do not impose non negativeness, in particular.
- The vector ζ_{t-1} is $\mathcal{F}_{t-1}$-measurable and updates the selected partial correlations at time t. Such ζ_{t-1} must be built so that $\mathbb{E}[\zeta_{t-1}] \simeq \mathbb{E}[Pc_{t-1}]$. This procedure is in line with usual updating equations in GARCH-type models. Recall that partial correlations are expectations of products of the two

different quantities $v_{k|L,t}$, for some $L \subset \{1,\dots,N\}$ and $k \notin L$, which are defined as

$$v_{k|L,t} = \frac{\epsilon_{k,t} - \mathbb{E}_{t-1}\left[\epsilon_{k,t}|\epsilon_{L,t}\right]}{\sqrt{h_{k|L,t}}},$$

where $\epsilon_{L,t} = \left(\epsilon_{i,t}\right)_{i\in L}$, and $\mathbb{E}_{t-1}\left[\epsilon_{k,t}|\epsilon_{L,t}\right]$ 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" $\epsilon_{k,t} - \mathbb{E}_{t-1}\left[\epsilon_{k,t}|\epsilon_{L,t}\right]$ is denoted by $h_{k|L,t}$. Then, by definition, we set $\zeta_{(ij|L),t-1} := v_{i|L,t-1}\, v_{j|L,t-1}$, for any edge $(ij|L)$.

- We apply a deterministic transformation Ψ to Pc_t. It twists the univariate dynamics to manage the constraint that partial correlations remain in $(-1,1)$. For the sake of simplicity, Ψ will be known.[2] To fix the ideas, Ψ is defined from $]-1,1[^{N(N-1)/2}$ to $\mathbb{R}^{N(N-1)/2}$ as

$$\Psi(Pc_t) = \left(\psi(\rho_{ij|L,t},\ (ij|L) \text{ are edges of } V(N))\right)',\ \psi(x) = \tan(\pi x/2).$$

  Alternatively, Ψ could be chosen among the sigmoïd functions for instance, for which $\psi(x) = (\exp(\alpha x) - 1)/(\exp(\alpha x) + 1)$ for some $\alpha \in \mathbb{R}$.

- The function $F_{\text{vine}}(\cdot)$ 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 et al. (2009). It is defined from $]-1,1[^{N(N-1)/2}$ to itself.

- vechof$(\cdot)$ denotes the operator "devectorization", that transforms a vector into a symmetric matrix. It maps a $N(N-1)/2$ column vector of correlations to a $N \times N$ correlation matrix.

By definition, equations (1)–(6) define a so-called vine-GARCH(p,q) model.[3]

Note that our vine-GARCH framework encompasses the constant conditional correlations case (the CCC model). Indeed, any regular vine introduces a family of partial correlations, whose constant values can be chosen to recover the constant correlation matrix (Theorem 1). In this case, simply set the coefficients Ξ_k and Λ_l to zero.

2.4. Parsimony

In full generality, this simplified version of the vine-GARCH(p, q) model still encompasses $(p+q+1)N(N-1)/2$ parameters. However, this approach can easily become more parsimonious and would provide a nice alternative to full DCC-GARCH models. Indeed, one can set constraints to any level of the tree (say r) by choosing zero partial correlations at and after the r-th tree in the underlying vine. We guess this should not modify significantly the (true) correlation dynamics, at least when r is large enough. This is due to the fact that partial correlations with nonempty conditioning subsets are correlations between residuals. In practice it is likely that these residuals tend to behave more and more as white noise when

the number of conditioning variables increases and for a well-chosen R-vine. By canceling partial correlations after the step r, we get a particular model with less parameters than in the full vine-GARCH specification. And whatever the chosen structure of the vine is after level r, the reconstruction formulas (3) provide the same correlation matrices. This is a nice theoretical property called $r - VF$ (see a formal definition in Section 2 in the appendix). A slightly different simplification of our vine-GARCH models would be to assume constant (non zero) partial correlations after some level (say r) in the vine. But in this case, we cannot ensure a similar r-VF property.

Remark 1. Obviously, alternative dynamics could extend our vine-GARCH(p,q) specification (5). For instance, it could be possible to modify the model to include nonlinear features such as asymmetries, switching regimes, time-varying parameters, exogenous variables, etc. A whole class of models is now open, based on partial correlations, exactly as the original GARCH framework of Bollerslev (1986) has been modified and revisited.

2.5. Vine Selection

The methodology above can be applied to any R-vine on N elements. Actually, the structure of the underlying R-vine may be seen as an additional parameter, independently of θ_c. 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, once some important factors have been found in the first trees.

To do so with a C-vine, we follow the sequential method developed by Dissmann, Brechmann, Czado, and Kurowicka (2013). This method consists in starting to compute the Kendall's tau of all the couples of nodes, and selecting the variable which induces the highest degree of dependence with the other ones. In the second tree, we compute a Kendall's tau per edge, but conditional on the variable chosen on the first tree. We use nonparametric statistics proposed by Veraverbeke, Omelka, and Gijbels (2011) to compute these quantities. We apply the same selection criteria to choose the convenient variable and proceed with the next trees similarly, until the last tree. The Kendall's tau is used as a dependence measure because it can easily be estimated, but other dependence measures are possible. This selection procedure is "bottom-up". Alternative methodologies exist, in particular the "top-down" procedure of Kurowicka (2011).

3. STATISTICAL INFERENCE BY QML

We can estimate vine-GARCH(p,q) models by maximizing a likelihood function that does not correspond to the true Data Generating Process necessarily, following the Quasi-Maximum Likelihood (QML) methodology, as explained in Gourieroux, Monfort, and Trognon (1984), Bollerslev and Wooldridge (1994) or White (1994), among others.

3.1. The QML Estimator

We choose a standard Gaussian QML estimator: we do a MLE as if (η_t) were a Gaussian white noise, but for inference purpose only. Obviously, the "true" underlying distributions of these innovations may be different. Note that the η_t-law can be estimated empirically a posteriori from a sample of residuals $R_t(\hat{\theta})^{-1/2}\epsilon_t$. We denote $\underline{\epsilon} = (\epsilon_1, \ldots, \epsilon_T)$ the vector of observations. Now using the assumed independence of the innovations η_t and developing H_t as $D_t R_t D_t$, the quasi-likelihood function of a path $(\epsilon_t)_{t=1,\ldots,T}$ is written as

$$L_T(\theta; \underline{\epsilon}) = \prod_{t=1}^{T} \exp\left\{-\frac{1}{2}\left(N\log(2\pi) + \log(|D_t R_t D_t|) + \epsilon_t' D_t^{-1} R_t^{-1} D_t^{-1} \epsilon_t\right)\right\}$$

$$= \prod_{t=1}^{T} \exp\left\{-\frac{1}{2}\left(N\log(2\pi) + \log\left(|D_t^2|\right) + \epsilon_t' D_t^{-2}\epsilon_t - u_t' u_t + \log(|R_t|) + u_t' R_t^{-1} u_t\right)\right\},$$

where $D_t = \operatorname{diag}\left(h_{1,t}^{1/2}, \ldots, h_{N,t}^{1/2}\right)$, and $u_t = \left(\epsilon_{1,t}/h_{1,t}^{1/2}, \ldots, \epsilon_{N,t}/h_{N,t}^{1/2}\right)' = D_t^{-1}\epsilon_t$ is the vector of GARCH standardized residuals. Thus, the quasi-log-likelihood function is the sum of two parts: the "variance part" of the likelihood, that depends on θ_v, and the "correlation part", that depends on both θ_v and θ_c. This is the usual decomposition of a variance and correlation part, as noticed in Engle (2002):

$$\ln L_T(\theta; \underline{\epsilon}) \propto \sum_{t=1}^{T}\left\{\log\left(|D_t^2|\right) + \epsilon_t' D_t^{-2}\epsilon_t - u_t' u_t + \log(|R_t|) + u_t' R_t^{-1} u_t\right\}$$

$$:= QL1_T(\theta_v; \underline{\epsilon}) + QL2_T(\theta_c, \theta_v; \underline{\epsilon}), \text{ where}$$

$$QL1_T(\theta_v; \underline{\epsilon}) := \sum_{t=1}^{T}\left\{\log\left(|D_t^2|\right) + \epsilon_t' D_t^{-2}\epsilon_t\right\} := \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\log\left(h_{i,t}\right) + \frac{\epsilon_{i,t}^2}{h_{i,t}}\right], \text{ and}$$

$$QL2_T(\theta_c, \theta_v; \underline{\epsilon}) := \sum_{t=1}^{T}\left\{\log(|R_t|) + u_t' R_t^{-1} u_t - u_t' u_t\right\}.$$

Therefore, we can first optimize $QL1_T$ over θ_v, and then over θ_c in $QL2_T$ conditionally on $\hat{\theta}_{T,v}$.

Our estimate $\hat{\theta}_{T,v}$ of θ_v is then

$$\hat{\theta}_{T,v} = \underset{\theta_v}{\arg\min}\ QL1_T(\theta_v; \underline{\epsilon}) := \sum_{i=1}^{N}\sum_{t=1}^{T}\left[\log\left(h_{i,t}\right) + \frac{\epsilon_{i,t}^2}{h_{i,t}}\right]. \tag{7}$$

The Newton–Raphson method is applied to solve this system. Note that $\hat{\theta}_{T,v}$ determines the (now estimated) variance processes $(h_{i,t})$ and then the (estimated) residuals u_t, denoted by $\hat{u}_t$. Given $\hat{\theta}_{T,v}$, a QML estimator of θ_c is obtained as

$$\hat{\theta}_{T,c} = \underset{\theta_c}{\arg\min}\ QL2_T\left(\theta_c, \hat{\theta}_{T,v}; \underline{\epsilon}\right) := \sum_{t=1}^{T}\left\{\log\left(|R_t|\right) + \hat{u}_t' R_t^{-1}\hat{u}_t\right\}. \tag{8}$$

Note that the calculation of R_t is obtained by the knowledge of Pc_t (the set of partial correlations that are known at $t-1$, by updating equation (5)) and by applying the previous function $F_{\text{vine}}(\cdot)$.

Strictly speaking, all the likelihood equations above depend on the initial values ϵ_0, D_0 and R_0. To fix the ideas, we propose to initialize them by their sample counterparts: for all $i = 1, \ldots, N$, set $\epsilon_0 = 0$, $\tilde{h}_{i,0} = \frac{1}{T-1}\sum_{t=1}^{T}\epsilon_{i,t}^2$, $\tilde{D}_0 = \text{diag}(\tilde{h}_{1,0}^{1/2}, \ldots, \tilde{h}_{N,0}^{1/2})$, and $\tilde{R}_0$ is the empirical correlation matrix of the sample path $(\epsilon_1, \ldots, \epsilon_T)$.

Remark 2. The absence of volatility spill-over effects allows for the estimation of θ_v through N simple optimizations independently. Obviously, if we remove this assumption, such an estimator can still be obtained by (7). But, in general, this would require an optimization in a high-dimensional space, a task that becomes harder and harder with increasing N.

Remark 3. It is possible to choose another QML parametric family that would be more adapted to fat tailed distributions typically (for instance the multivariate Student law, or any elliptical distribution). But then, we would lose the nice property of a two-stage estimation procedure, that is so important in practice.

3.2. Estimation of Innovations ζ_t

Now, let us focus on the estimation of the (R_t)-dynamics.

At time t, the vector ζ_t is a key information as it drives the shocks on the partial correlation processes. Since the variables $v_{k|L,t}$ are not observable, this is the case for ζ_t too. Nonetheless, we can replace the latter quantities by some estimates. Indeed, we can evaluate $\mathbb{E}_{t-1}\left[\epsilon_{k,t}|\epsilon_{L,t}\right]$ and $h_{k|L,t}$ to get $\hat{v}_{k|L,t}$, an approximated value of $v_{k|L,t}$. Then, *by construction*, the $N(N-1)/2$-sized approximated vector ζ_t, now denoted by $\hat{\zeta}_t$, will stack 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 $\hat{\zeta}_t$ will be the same as for Pc_t.

Here, we propose two ways for evaluating $v_{k|L,t}$, and then ζ_t.

The first method is "ad-hoc" and is based on linear regressions of $\epsilon_{k,t}$ on $\epsilon_{L,t}$:

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

for some time-varying parameters $\alpha_{k|L,t}$ and $\beta_{k|L,t}$. They will be estimated through "rolling window" OLS. When they are calculated with the m vectors ϵ_{t-k}, $k = 1, \ldots, m$, for some windows size m, they are denoted by $\hat{\alpha}_{k|L,t}$ and $\hat{\beta}_{k|L,t}$ and are $\mathcal{F}_{t-1}$-measurable. Then, we approximate $\epsilon_{k,t} - \mathbb{E}[\epsilon_{k,t}|\epsilon_{L,t}]$ by $\epsilon_{k,t} - \hat{\alpha}_{k|L,t} - \hat{\beta}_{k|L,t}'\epsilon_{L,t}$ and an empirical "rolling-window" estimator of $h_{k|L,t}$

is defined by $\hat{h}_{k|L,t} := m^{-1}\sum_{i=1}^{m}(\epsilon_{k,t-i} - \hat{\alpha}_{k|L,t-i} - \hat{\beta}'_{k|L,t-i}\epsilon_{L,t-i})^2$. The size m should increase with T in theory but trying to exhibit some "optimal" m, for instance by some cross-validation techniques, is beyond the scope of the present work. We get $\hat{\upsilon}_{k|L,t} = (\hat{\alpha}_{k|L,t} + \hat{\beta}'_{k|L,t}\epsilon_{L,t})/\sqrt{\hat{h}_{k|L,t}}$, and then $\hat{\zeta}_t$, that will replace ζ_t is (5). 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.

The second method is based on the theoretical distribution of the residuals ϵ_t given $\mathcal{F}_{t-1}$, that is unknown at this stage. In accordance with our Gaussian QMLE, we assume that the latter distribution is elliptical. Then, its first two conditional moments are calculated easily. Indeed, if a vector $(X, Y)'$ is elliptical with $\Sigma_{XX} = \mathrm{Var}(X)$, $\Sigma_{YY} = \mathrm{Var}(Y)$, $\Sigma_{XY} = \Sigma'_{YX} = \mathrm{Cov}(X, Y)$, then $\mathbb{E}[X|Y] = \mathbb{E}[X] + \Sigma_{XY}\Sigma_{YY}^{-1}(Y - \mathbb{E}[Y])$ and $\mathrm{Var}(X|Y) = \Sigma_{XX} - \Sigma_{XY}\Sigma_{YY}^{-1}\Sigma_{YX}$: see Corollary 5 in Cambanis, Huang, and Simons (1981). Hence we can calculate $\hat{\upsilon}_{k|L,t} = (\epsilon_{k,t} - \mathbb{E}_{t-1}[\epsilon_{k,t}|\epsilon_{L,t}])/\sqrt{h_{k|L,t}}$. To be specific, under these assumptions, we write

$$\mathbb{E}_{t-1}[\epsilon_{k,t}|\epsilon_{L,t}] = \mathrm{Cov}_{t-1}(\epsilon_{k,t}, \epsilon_{L,t})\,\mathrm{Var}_{t-1}(\epsilon_{L,t})^{-1}\epsilon_{L,t},$$

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

and the latter conditional covariances are $\mathcal{F}_{t-1}$ measurable, i.e., are known at t. In the empirical part (Section 4), this second method of calculation of $\hat{\zeta}_t$ is used for convenience (higher computation speed).

3.3. Estimation Strategy

Unfortunately, the underlying process (R_t) induces tricky computations of scores and Hessians for $QL2_T$. This is the case for both DCC and vine Garch dynamics. Here, we propose two strategies depending on the dimensionality of the problem.

As the general DCC model, the vine GARCH specification may suffer from the curse of dimensionality. However, when the matrices of parameters Ξ_j and Λ_k are diagonal (a usual situation), it is possible to weaken drastically this problem by proceeding sequentially. Indeed, in partial correlation R-vines, any partial correlation on tree T_k can be updated (through the ζ_t quantities or rather their estimates $\hat{\zeta}_t$ in practice) easily knowing the partial correlations on the previous trees $T_{k'}$, $k' \leq k-1$.[4]

Let us detail the sequential procedure for a C-vine, w.l.o.g. Instead of relying on a brute-force optimization in high dimension, a vine-GARCH model based on a C-vine may be estimated by solving $N \times (N-1)/2$ simple optimization programs, related to the bivariate dynamics that are associated with any edge. This means we estimate successively the dynamics of $(\epsilon_{i,t}, \epsilon_{j,t})$ where the $N \times$

$(N-1)/2$ couples (i,j) describe the conditioned subsets of all the edges in the underlying C-vine, starting from the bottom tree.

To be even more explicit, denote the edges of the vine by $\{(ij|L)\}$, and the unknown matrix parameters as $\Omega := [\omega_{(ij|L)}]$, $\Xi_k := \text{diag}(\xi_{(ij|L);k})$, $k = 1, \ldots, p$ and $\Lambda_l := \text{diag}(\lambda_{(ij|L);l})$, $l = 1, \ldots, q$. Assume the underlying C-vine is given in Figure 1. In particular, 1 is the root in the first tree. The $N-1$ first partial correlation dynamics are "usual" correlation processes and depend on the estimated volatility and the observations. The parameters of these $N-1$ first processes can be minimized *independently* based on the objective functions

$$QL2_T^{1j}\left(\theta_{c,1j}, \hat{\theta}_{T,v}; \underline{\epsilon}\right) = \sum_{t=1}^{T}\left[\log|R_{(1j),t}| + \hat{u}'_{(1j),t}\left(R_{(1j),t}\right)^{-1}\hat{u}_{(1j),t}\right],\ j = 2, \ldots, N,$$

where $\hat{u}_{(1j),t} = [\hat{u}_{1,t}, \hat{u}_{j,t}]'$ and $R_{(1j),t}$ is the 2×2 correlation matrix of $(\epsilon_{1,t}, \epsilon_{j,t})$ given $\mathcal{F}_{t-1}$. With obvious notations, $\theta_{c,1j} = \left(\omega_{1j}, \boldsymbol{\xi}_{1j}, \boldsymbol{\lambda}_{1j}\right)$ are the remaining unknown parameters that are associated with the bivariate process $(\epsilon_{1,t}, \epsilon_{j,t})$.

Now, after conditioning on 1, there are $N-2$ dynamic partial correlations in T_2. Due to (11), they follow the ARMA-type dynamics

$$\psi\left(\rho_{2j|1,t}\right) = \omega_{2j|1} + \sum_{k=1}^{p}\xi_{2j|1;k}\,\psi\left(\rho_{2j|1,t-k}\right) + \sum_{l=1}^{q}\lambda_{2j|1;l}\,\hat{v}_{2|1,t-l}\,\hat{v}_{j|1,t-l},\ j = 3, \ldots, N.$$

For QML inference purpose, we assumed $\epsilon_t|\mathcal{F}_{t-1} \sim \mathcal{N}(0, H_t)$. As explained in Section 2.3, $\hat{v}_{k|1,t-1}$ above depends on the volatility processes, the observations and the correlations calculated from tree T_1. Hence, we estimate the partial correlations dynamics on tree T_2 by maximizing $N-2$ objective functions independently over each correlation parameter space of tree T_2, given the estimated correlations on T_1. The objective functions are, for all $j = 3, \ldots, N$

$$QL2_T^{2j|1}\left(\theta_{c,2j|1}, \hat{\theta}_{T,v}; \underline{\epsilon}, \hat{\rho}_{12}, \hat{\rho}_{1j}\right) = \sum_{t=1}^{T}\left[\log|R_{(2j),t}| + \hat{u}'_{(2j),t}\left(R_{(2j),t}\right)^{-1}\hat{u}_{(2j),t}\right].$$

Here, $R_{(2j),t}$ is the correlation matrix of $(\epsilon_{2,t}, \epsilon_{j,t})$ given $\mathcal{F}_{t-1}$. Its coefficient $\rho_{2j,t}$ is computed from the estimated dynamic partial correlations $\hat{\rho}_{2j|1,t}$ and the (estimated) correlations $\hat{\rho}_{1l,t}$, $l = 2, \ldots, N$. Obviously, $\hat{u}_{(j,k),t} = [\hat{u}_{j,t}, \hat{u}_{k,t}]'$.

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 estimation of any partial correlation process of a tree T_k depends only on a subset of partial correlations associated with the edges of T_{k-1} and before, invoking the recursive formula (3). Consider any edge $(ij|L)$ in T_k and denote by $\theta_{c,ij|L} = \left(\omega_{ij|L}, \boldsymbol{\xi}_{ij|L}, \boldsymbol{\lambda}_{ij|L}\right)$ the associated subvector of θ_c. For instance, with our C-vine of Figure 1, L does not depend on the conditioned subsets and is $L := L_i = \{1, \ldots, i-1\}$, $k = 2, \ldots, N-1$. Our iterative algorithm is summarized as

$$\hat{\theta}_{T,c,ij|L_i} = \underset{\theta_{c,ij|L_i}}{\arg\min}\, QL2_T^{ij|L_i}\left(\theta_{c,ij|L_i}, \hat{\theta}_{T,v}; \underline{\epsilon}, \hat{\rho}_{i-1,i|L_{i-1}}, \hat{\rho}_{i-1,j|L_{i-1}}\right),$$

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

We denote this strategy *C-vine* (*D-vine, or even R-vine*) *iterative process*, which is particularly effective when N becomes "large" (say larger than 5 assets). At each edge on a specific level, only $(p+q+1)$ parameters need to be estimated. Consequently, we also use the Sequential Quadratic Programming method when estimating the C-vine iterative process.

A drawback of the latter iterative process may be the propagation 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 ($N \leq 5$) to avoid this iterative method. But the nonlinearity and the instability of the likelihood function in the vine-GARCH case require another approach to maximize $QL2_T$. In such a case, we propose to use a stochastic algorithm, simulated annealing, that helps to prevent falling into local maxima. Note that the simulated annealing algorithm can also be used when estimating the model through the previous 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.

The asymptotic properties of the previous two-step quasi maximum likelihood estimator (consistency, asymptotic normality) have been proven, when it is assumed that any conditional variance series follows the univariate GARCH processes

$$h_{i,t} = \varsigma_i + \sum_{k=1}^{q_i} \kappa_{i,k}\epsilon_{i,t-k}^2 + \sum_{l=1}^{p_i} \tau_{i,l} h_{i,t-l}, \tag{9}$$

such that $\theta_v^{(i)} = (\varsigma_i, \boldsymbol{\kappa}_i, \boldsymbol{\tau}_i)' \in \mathbb{R}_+^{p_i+q_i+1}$ for all $i = 1,\ldots,N$. The required technical conditions overlap some usual ones in the M-GARCH world, as those in Engle and Sheppard (2001) or Francq and Zakoïan (2004). Nonetheless, the nonlinear feature of $F_{\text{vine}}(\cdot)$ induces particular difficulties. These results are available in Poignard and Fermanian (2016a), the extended working paper version of this article. Moreover, some sufficient conditions to obtain strongly stationary solutions of (6) are provided in Poignard and Fermanian (2016b) when $p = q = 1$.

4. EMPIRICAL APPLICATIONS

To simplify and to lighten notations, we restrict ourselves to first order models in this section. Moreover, we consider no cross-effects between all the individual partial correlation processes, i.e., the matrices Ξ_k and Λ_l are assumed to be diagonal. Then, when $p = q = 1$, the $N-1$ first elements of Pc_t correspond to usual correlations, i.e., $\rho_{ij|\emptyset,t} = \rho_{ij,t}$, and they follow the processes

$$\psi\left(\rho_{ij,t}\right) = \omega_{ij} + \xi_{ij}\,\psi\left(\rho_{ij,t-1}\right) + \lambda_{ij}\,\hat{v}_{i,t-1}\,\hat{v}_{j,t-1}, \tag{10}$$

with $\hat{v}_{k,t} = \epsilon_{k,t}/\sqrt{\hat{h}_{k,t}}$. From the N-th component on, the elements of Pc_t are "true" partial correlations for which $L \neq \emptyset$. Their dynamics are given by

$$\psi(\rho_{ij|L,t}) = \omega_{ij|L} + \xi_{ij|L}\,\psi(\rho_{ij|L,t-1}) + \lambda_{ij|L}\,\hat{v}_{i|L,t-1}\,\hat{v}_{j|L,t-1}. \tag{11}$$

Strictly speaking, the partial correlation dynamics we invoke for inference or simulation purpose are given by (10) and (11).

4.1. A Simulation Study

We consider multivariate series (ϵ_t) of size $N = 2, 6, 10, 20, 30, 50$ as data generating processes (DGP). Their innovations η_t are standardized normal white noises. The conditional covariance matrices of these processes are deduced from a MGARCH form $H_t = D_t R_t D_t$. To generate N univariate variance processes along (9), we randomly choose the corresponding $3N$ parameters. Denoting by $\mathcal{U}(a, b)$ the uniform distribution on $[a, b]$, we simulate $\varsigma \sim \mathcal{U}\left(10^{-5}, 9.10^{-5}\right)$, $\kappa \sim \mathcal{U}(0.01, 0.15)$ and $\tau \sim \mathcal{U}(0.85, 0.95)$, under the stationarity constraint $\kappa + \tau < 1$. As for the correlation dynamics, we first randomly choose $N(N-1)/2$ deterministic processes among the cosinus, sinus, modulo, and constant functions, and then generate some series

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

for every $t = 1, \ldots, T$. Our parameters $a_1, a_2, b_1, b_2, c_1, c_2, d_1, d_2$ are chosen randomly and independently following a $\mathcal{U}(-0.4, 0.4)$ and α, β, μ are randomly (equally) selected among the fixed subset $\{100, 200, 500, 1,000, 1,500, 2,000\}$. All these series constitute the components of a lower triangular matrix K_t with ones on the main diagonal. Then, we generate symmetric and positive definite matrices $C_t = K_t K_t'$ and $R_t = C_t^{\star -1/2} C_t C_t^{\star -1/2}$, where $C_t = [c_{ij,t}]$ and $C_t^{\star} = \mathrm{diag}(c_{11,t}, \ldots, c_{NN,t})$. 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 randomly and given ϵ_1, we simulate the successive values of a MGARCH process with conditional covariance matrices (H_t). We do this iterative procedure for $T = 10,000$ 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, a scalar DCC and two GAS-type dynamics derived from Creal et al. (2011), the hyperspherical GAS and the projected GAS (see Section 1 in the appendix). In this study, the DCC and the projected GAS specification are not highly parameterized in terms of correlation dynamics: the scalar DCC (resp. diagonal QFDCC, resp. projected GAS) requires the estimation of 3 (resp. $3N$, resp. 3) parameters, after correlation targeting. However, the hyperspherical GAS requires the estimation of $2 + N(N-1)/2$ parameters due to the estimation of a vectorial intercept parameter. We thus estimated the hyperspherical GAS for

$N \in \{2, 6, 10\}$. The Sequential Quadratic Programming method has been implemented to estimate these dynamical models, since it is well-suited for constrained optimization with a "reasonable" number of parameters. As a benchmark, we also compute the empirical correlation matrices of our returns through a rolling-window of size 200 observations.[5] Some details about practical implementation are given in Section 3 in the appendix.

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), $\hat{R}_t^{rw}$ (rolling-window), $\hat{R}_t^{pgas}$ (projected GAS) and $\hat{R}_t^{hgas}$ (hyperspherical GAS). Moreover, we consider a constrained version of the vine-GARCH. For $N = 6, 10$, the partial correlations of the last two trees are constrained to their unconditional partial correlation values, as estimated over the whole sample. For $N = 20$, the partial correlations from the 11th level are set to their unconditional partial correlation values. The same applies for $N = 30, 50$, where we only consider the constrained vine-GARCH case. Alternatively, we could set zero partial correlations for these two last trees of the C-vine and the results would be comparable. We denote by $\hat{R}_t^{vine\star}$ 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 of Section 2.5.

We compare the true correlation process and the estimated correlation 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 compute the previous norm for each t and for $A = R_t$ and

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

We take the average of those quantities over $T = 10,000$ periods of time. Since we repeat this experiment 300 times, this provides an average gap for all those simulations. Table 1 reports the results.

The C-vine model (constrained or not) clearly outperforms the other specifications, except when the portfolio size is very small. In the latter case ($N = 2$, here), the GAS-type models and the QFDCC are the best ones. But as soon as N becomes larger than 5, both the DCC-type models and the projected GAS display a significant gap, which highlights that they are not sufficiently flexible to capture

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

$\|\|R_t - B\|\|_F$	$B = \hat{R}_t^{dcc}$	$B = \hat{R}_t^{qfdcc}$	$B = \hat{R}_t^{rw}$	$B = \hat{R}_t^{pgas}$	$B = \hat{R}_t^{hgas}$	$B = \hat{R}_t^{vine}$	$B = \hat{R}_t^{vine\star}$
$N = 2$	0.0397	0.0378	0.0829	0.0390	0.0388	0.0407	-
$N = 6$	0.4995	0.4791	0.5275	0.5012	0.4789	0.3906	0.4137
$N = 10$	0.8270	0.9237	0.8784	0.8954	1.0358	0.6413	0.6825
$N = 20$	1.6931	2.0106	1.7372	-	-	1.3250	1.3766
$N = 30$	2.4876	2.6681	2.5151	-	-	-	2.0583
$N = 50$	3.2839	3.8662	3.7691	-	-	-	2.6800

complex dynamics of correlation matrices with only two parameters. Surprisingly, the hyperspherical GAS model does not seem to do a better job, while it depends on $2 + N(N-1)/2$ parameters for correlations. This number of free parameters is even larger in the vine-GARCH case ($3N(N-1)/2$), inducing even more flexibility than hyperspherical GAS models. Apparently, this matters in practice. Note that the constrainted version of the vine-GARCH model yields good results with a reduced set of free parameters (its size is $O(N)$). As for the rolling-window correlation, the result emphasizes this empirical measure should be taken with care. The rolling nature of the samples makes the rolling-window correlation very slow to react to a rapid correlation fluctuation.

4.2. Discussion

In addition to the previous empirical results, additional elements argue in favor of vine-GARCH specifications. Indeed, the general form of GAS models shares the same drawback as the DCC one when the dimension grows. The curse of dimensionality hampers any high-dimension modelling, which implies reduced form dynamics such as the scalar/diagonal DCC or GAS. By contrast, the vine-GARCH model allows for flexible and richly parameterized dynamics, while one can carry out the iterative procedure and estimate 3 parameters for each partial correlation process. The dimension reduction is strongly fostered when applying the truncation procedure. Both DCC and GAS specification do not allow for such iterative methods so that one must estimate all the parameters at once.

Moreover, if one considers the GAS model specified as given by equations (15) and (16) in Creal et al. (2011), Section 3.1, there is no guarantee that the induced correlation matrices are positive definite. In general, this necessitates the use of projection methods as above with the projected GAS model. Alternatively, the hyperspherical model ensures positive-definiteness, but the associated parameters (angles) are not interpretable at all. By contrast, vine-GARCH models provide dynamics that can be easily interpreted. They automatically generate "true" positive definite correlation matrices. They can be estimated in large dimension iteratively and, after parallelization possibly (see 3).

4.3. Application to Real Portfolios

In this subsection, we estimate by Quasi-Maximum Likelihood the DCC-GARCH and vine-GARCH models for two financial portfolios. They are composed of a daily series of stock log-returns related to the Morgan Stanley Capital International (MSCI) Developed Markets indices. In the so-called 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 2013, which amounts to 3,669 observations.

TABLE 2. GARCH(1,1) Models estimated by QML for 9 stock indices. The Bollerslev–Wooldridge standard deviations are in parentheses

Asset	ς	κ	τ
Australia	0.657e-5 (0.114e-5)	0.124 (0.014)	0.846 (0.011)
France	0.388e-5 (0.076e-5)	0.111 (0.009)	0.876 (0.008)
Germany	0.368e-5 (0.080e-5)	0.100 (0.011)	0.889 (0.010)
Greece	0.191e-5 (0.147e-5)	0.090 (0.010)	0.917 (0.015)
Italy	0.235e-5 (0.052e-5)	0.113 (0.010)	0.883 (0.008)
Japan	0.997e-5 (0.157e-5)	0.103 (0.012)	0.849 (0.013)
Netherlands	0.363e-5 (0.069e-5)	0.110 (0.010)	0.876 (0.009)
United Kingdom	0.338e-5 (0.067e-5)	0.115 (0.011)	0.868 (0.009)
United States	0.223e-5 (0.056e-5)	0.102 (0.010)	0.884 (0.008)

First, we have centered the time series by assuming that $\mathbb{E}_{t-1}[r_t] = \mu_t(\theta)$ follows a one-order autoregressive process (estimated by OLS). Second, we estimate the conditional variance processes of the components of $\epsilon_t = r_t - \mu_t$. The GARCH(1,1) specification was chosen a priori for modelling these marginal dynamics. Indeed, this is by far the main reference model used in the literature. The estimation results are reported in Table 2.

We now turn to the second QML step, i.e., the estimation of the conditional correlation dynamics, knowing the GARCH(1,1) estimates. For portfolios I and

TABLE 3. C-vine-GARCH estimated by QML for Portfolio I: Germany (1), United Kingdom (2), Italy (3), France (4), and Netherlands (5). The Bollerslev–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 4. Scalar DCC-GARCH estimated by QML for portfolio I. The Bollerslev–Wooldridge standard deviations are in parentheses

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

TABLE 5. Diagonal QFDCC estimated by QML for Portfolio I. The Bollerslev–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)

II, we select a relevant C-vine, according to the Kendall's tau selection procedure (see Section 2.5). We associate an index with each country. This number corresponds to the index of the tree for which this country is the "center" (the node with maximal connections). Since Portfolio I is composed of European stocks, it can be considered as relatively homogenous, including the main countries of the Eurozone. The selecting procedure induces 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 correlations of two countries, given Germany, on Tree 2. Then, on Tree 3, the conditioning subset is Germany (1) and United Kingdom (2), etc. The composition of the "heterogenous" portfolio II is given as follows: Germany (1), Greece (2), United States (3), Italy (4), Japan (5), and Australia (6).

Actually, we consider two cases of C-vine-GARCH models. The first one is the usual unconstrained C-vine tree. The second one is a constrained version of the previous one, where the partial correlations of the last two 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. In every case, the parameters are estimated by simulated annealing. Table 3 reports the estimation

TABLE 6. C-vine-GARCH Model estimated by QML for Portfolio I: Germany (1), United Kingdom (2), Italy (3), France (4), and Netherlands (5). The Bollerslev–Wooldridge standard deviations are in parentheses

Ω	Estimate (Std Err)	Ξ	Estimate (Std Err)	Λ	Estimate (Std Err)
ω_{12}	−0.0661 (0.0174)	ξ_{12}	0.9769 (0.0433)	λ_{12}	0.1932 (0.0193)
ω_{13}	−0.0771 (0.0441)	ξ_{13}	0.9804 (0.0659)	λ_{13}	0.1986 (0.0182)
ω_{14}	−0.1665 (0.6173)	ξ_{14}	0.9923 (0.1121)	λ_{14}	0.2590 (0.0638)
ω_{15}	−0.0858 (0.0709)	ξ_{15}	0.9915 (0.0613)	λ_{15}	0.1554 (0.0431)
$\omega_{23\|1}$	0.0081 (0.0047)	$\xi_{23\|1}$	0.9799 (0.1265)	$\lambda_{23\|1}$	0.0013 (0.0165)
$\omega_{24\|1}$	0.0248 (0.0666)	$\xi_{24\|1}$	0.9577 (0.0934)	$\lambda_{24\|1}$	0.0112 (0.0113)
$\omega_{25\|1}$	0.0172 (0.0081)	$\xi_{25\|1}$	0.9641 (0.0329)	$\lambda_{25\|1}$	0.0135 (0.0221)
$\omega_{34\|12}$	1.0821	$\xi_{34\|12}$	-	$\lambda_{34\|12}$	-
$\omega_{35\|12}$	0.6300	$\xi_{35\|12}$	-	$\lambda_{35\|12}$	-
$\omega_{45\|123}$	0.7957	$\xi_{45\|123}$	-	$\lambda_{45\|123}$	-

TABLE 7. C-vine-GARCH estimated by QML for Portfolio II: Germany (1), Greece (2), United States (3), Italy (4), Japan (5), and Australia (6). The Bollerslev–Wooldridge standard deviations are in parentheses

Ω	Estimate (StdE)	Ξ	Estimate (StdE)	Λ	Estimate (StdE)
ω_{12}	−0.0008 (0.0359)	ξ_{12}	0.9823 (0.1935)	λ_{12}	0.0387 (0.0227)
ω_{13}	0.0016 (0.0036)	ξ_{13}	0.9821 (0.0530)	λ_{13}	0.0382 (0.0052)
ω_{14}	−0.0694 (0.0259)	ξ_{14}	0.9801 (0.0180)	λ_{14}	0.1915 (0.0092)
ω_{15}	0.0046 (0.0081)	ξ_{15}	0.9777 (0.0265)	λ_{15}	0.0146 (0.0043)
ω_{16}	0.0017 (0.0058)	ξ_{16}	0.9835 (0.0142)	λ_{16}	0.0288 (0.0012)
$\omega_{23\|1}$	−0.0072 (0.0399)	$\xi_{23\|1}$	0.9334 (0.4570)	$\lambda_{23\|1}$	−0.0007 (0.0037)
$\omega_{24\|1}$	0.0043 (0.0156)	$\xi_{24\|1}$	0.9837 (0.2434)	$\lambda_{24\|1}$	0.0001 (0.0100)
$\omega_{25\|1}$	0.0129 (0.0272)	$\xi_{25\|1}$	0.9384 (0.0790)	$\lambda_{25\|1}$	0.0028 (0.0061)
$\omega_{26\|1}$	0.0022 (0.0076)	$\xi_{26\|1}$	0.9906 (0.0238)	$\lambda_{26\|1}$	0.0049 (0.0175)
$\omega_{34\|12}$	0.0009 (0.0013)	$\xi_{34\|12}$	0.9729 (0.0187)	$\lambda_{34\|12}$	0.0107 (0.0082)
$\omega_{35\|12}$	−0.0001 (0.0002)	$\xi_{35\|12}$	0.9953 (0.0045)	$\lambda_{35\|12}$	0.0047 (0.0035)
$\omega_{36\|12}$	−0.0004 (0019)	$\xi_{36\|12}$	0.9888 (0.0393)	$\lambda_{36\|12}$	0.0053 (0.0118)
$\omega_{45\|123}$	0.0311	$\xi_{45\|123}$	-	$\lambda_{45\|123}$	-
$\omega_{46\|123}$	0.2472	$\xi_{46\|123}$	-	$\lambda_{46\|123}$	-
$\omega_{56\|1234}$	0.8669	$\xi_{56\|1234}$	-	$\lambda_{56\|1234}$	-

results of the vine-GARCH model for the unconstrained case. For the sake of comparison, Table 4 (resp. Table 5) provides the estimation results of the scalar DCC (resp. diagonal QFDCC). The results for the constrained case are very close to those of the unconstrained case: see Tables 6 and 7.

The same model is implemented for portfolio II, which is heterogenous in terms of geographical areas. Table 8 (resp. Table 9, Table 10) reports the estimation results of the C-vine-GARCH (resp. diagonal QFDCC, scalar DCC).

TABLE 8. Vine-GARCH estimated by QML for Portfolio II: Germany (1), Greece (2), United States (3), Italy (4), Japan (5), and Australia (6). The Bollerslev–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 9. Diagonal QFDCC estimated by QML for Portfolio II. The Bollerslev–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)

Concerning Portfolio I, the higher the level of the tree is, the smaller are the partial correlation coefficients ω and λ. We may infer that once we control for the information given by Germany (1) (the core of the Eurozone) and United Kingdom (2), the dynamics of partial correlations on trees T_3 and T_4 are not very informative. This looks like evaluating a white noise. This is confirmed by the modelling of constrained vines, where the 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 United States 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.

4.4. Specification Testing

Once the model is estimated, we are able to forecast the covariance matrices 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 (2009) for a survey. In this study, we focus on direct out-of-sample evaluation methods, which allow for pairwise comparisons. They test whether some of the previous models provide better forecasts in terms of portfolio volatility behavior. Following the methodology of Engle and Colacito (2006), we develop a mean-variance portfolio approach to test the H_t forecasts. Intuitively, if a conditional covariance process is misspecified, then the minimum variance portfolio should emphasize such a shortcoming, compared to other models. Then, consider an investor who allocates a fixed amount between N stocks, according to a minimum-variance strategy and independently at each time t. At each date t, he/she solves

TABLE 10. Scalar DCC GARCH estimated by QML for Portfolio II. The Bollerslev–Wooldridge standard deviations are in parentheses

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

TABLE 11. Diebold Mariano Test of Multivariate GARCH models for Portfolio I

	DCC	QFDCC	O-GARCH	C-vine	C-vine-c
DCC		0.6509	−5.9350***	0.7784	0.3551
QFDCC	−0.6509		−6.1426***	0.4237	0.0475
O-GARCH	5.9350***	6.1426***		5.9438***	5.6779***
C-vine	−0.7784	−0.4237	−5.9498***		2.1206**
C-vine-c	−0.3551	−0.0475	−5.6779***	−2.1206**	

Rejection of the null hypothesis at: 10% for *, 5% for **, 1% for ***. When the null hypothesis of equal predictive accuracy is rejected, a positive number is evidence in favor of the model in the column.

$$\min_{w_t} w_t' H_t w_t, \text{ s.t. } \iota' w_t = 1, \tag{12}$$

where w_t is the $N \times 1$ vector of portfolio weights chosen at (the end of) time $t-1$, ι is a $N \times 1$ vector of 1 and H_t is the estimated conditional covariance matrix of the asset returns at time t. They are deduced from some dynamics that have been estimated on the subsample January 1999–October 2011. Once the latter process is estimated in-sample, out-of-sample predictions are plugged into the program (12) between November 2011 and August 2013. The solution of (12) is given by the global minimum variance portfolio $w_t = H_t^{-1}\iota/\iota' H_t^{-1}\iota$.

Engle and Colacito (2006) show that the realized portfolio volatility is the smallest one when the model covariance matrices are correctly specified. As a consequence, if wealth is allocated using two different dynamic models i and j, whose predicted 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, defined as $u_{ij,t} = \left\{w_{i,t}'\epsilon_t\right\}^2 - \left\{w_{j,t}'\epsilon_t\right\}^2$. The portfolio variances are the same if the predicted covariance matrices are the

TABLE 12. Diebold Mariano Test of Multivariate GARCH models for Portfolio II

	DCC	QFDCC	O-GARCH	C-vine	C-vine-c
DCC		−0.6220	−4.9369***	0.0908	−0.7952
QFDCC	0.6220		−4.9783***	0.2650	−0.5991
O-GARCH	4.9369***	4.9783***		4.6416***	4.1741***
C-vine	−0.0908	−0.2650	−4.6416***		−3.0709***
C-vine-c	0.7952	0.5991	−4.1741***	3.0709***	

Rejection of the null hypothesis at: 10% for *, 5% for **, 1% for ***. When the null hypothesis of equal predictive accuracy is rejected, a positive number is evidence in favor of the model in the column.

same. Thus we test the null hypothesis $\mathcal{H}_0 : \mathbb{E}\left[u_{ij,t}\right] = 0$ by the Diebold and Mariano (1995) test. It consists of a least square regression using HAC standard errors, given by $u_{ij,t} = \alpha + \epsilon_{u,t}$, $\mathbb{E}[\epsilon_{u,t}] = 0$, and we test $\mathcal{H}_0 : \alpha = 0$. If the mean of $u_{ij,t}$ is significantly positive (resp. negative), then the forecasts given by the covariance matrices of model j (resp. i) are preferred.

We run the latter test for portfolios I and II and to compare the scalar DCC, QFDCC, constrained C-vine-GARCH (C-vine-c) and unconstrained C-vine-GARCH (C-vine) models. We also compare these parameterizations to a factor model, the O-GARCH(1,1).[6] The results are reported in Tables 11 and 12. Those tables provide the out-of-sample Dielbold–Mariano test statistics that check the equality of a pair of series of covariance matrices using the loss function $u_{ij,t}$ over the period November 2011–August 2013.

We first note that in the homogenous case, the DCC specifications do not provide better 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 United States, there remains a significant amount of idiosyncratic risk. Both versions of the C-vine are not outperformed by the scalar DCC, and the C-vine provides better covariance forecasts than the QFDCC. The QFDCC is also slightly outperformed by the scalar DCC specification for the heterogenous portfolio, which is rather surprising. Finally, the O-GARCH model is beaten by all the others distinctly. But all these results are not sufficiently clear-cut to draw any strong conclusion concerning a potential hierarchy between all these models, at least in terms of a "naive" investment strategy.

5. CONCLUSION

We have proposed to rely on vines to define a new family of multivariate GARCH-type models. The main feature of our methodology is the specification and the estimation of partial correlation processes "independently" and largely arbitrarily, and their use to generate sequences of correlation matrices in a very flexible and general way. The canonical vine is particularly intuitive to model a hierarchy between asset returns, as our reasonings are close to factor models. Besides, the vine-GARCH approach allows for building parsimonious models. Indeed, we can assume (theoretically and often empirically) no partial correlation dynamics (or at least, constant, simpler, homogeneous, etc., dynamics) at all edges in the vine from some level on. All these elements foster flexibility and enable to generate high-dimensional matrices.

Our vine-GARCH approach is apparently more able to reproduce heterogeneous and complex time-varying patterns of correlations than most standard competitors. This means our model is richer and more flexible than the latter ones. This is not a surprise because we are able to efficiently manage (equation-by-

equation) numerous parameters that have a direct effect on correlation dynamics. Nonetheless, different vines can induce different time series of correlation. The choice of the "best" vine for our purpose is a difficult question that will need further investigation.

Therefore, a new framework has been opened in the field of MGARCH models. Sufficient conditions for the consistency and the asymptotic normality of a two-step quasi-maximum estimator are provided in a companion working paper. 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 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 but more empirical work is probably necessary to evaluate all the advantages of such approaches w.r.t. more classic ones.

NOTES

1. This article corresponds to the merger of two working articles: Poignard and Fermanian (2016a) presents the vine-GARCH specification, states the weak consistency and the asymptotic normality of the estimates, and provides some empirical results; Poignard and Fermanian (2016b) provides the conditions for the existence and the uniqueness of stationary solutions of the vine-GARCH models.

2. There is no doubt the methodology could be extended to deal with a parametric function Ψ, i.e., that would depend on an unknown finite-dimensional additional parameter. Nonetheless, this would complicate inference.

3. Recently, we have been aware of So and Yeung (2014), who developed similar ideas independently. They proposed some time-varying parametric dynamics in R-vine models. Our approach is different regarding some key points: their estimators depend on a fully specified parametric copula model while ours are semiparametric since we specify only conditional second-order moments. Moreover, their copula parameters are updated through ad-hoc rolling window techniques, whereas we estimate the innovations ζ_t using the regression properties of partial correlations. Finally, no asymptotic theory is provided in So and Yeung (2014).

4. To be specific, at a given edge $(ij|L)$ of an R-vine, we need to calculate the $\zeta_{(ij|L),t-l}$, $l = 1, \dots, q$. As explained in 2.3, this necessitates the calculation of conditional covariances of the subvectors associated with the indices (iL) and (jL). Since L is the conditioning subset at this edge, this is always possible once we have evaluated all edge dynamics associated with the previous trees.

5. Some tests with $m = 50$ and $m = 100$ have yielded comparable results.

6. The O-GARCH assumes the decomposition $H_t = P\Lambda_t P'$, where $\Lambda_t = \mathrm{diag}(\lambda_{1,t}, \dots, \lambda_{K,t})$, with K the number of factors. Here, we choose $K = N$ factors and each λ_t is supposed to follow a univariate GARCH(1,1) process that is estimated by maximum likelihood. The matrix P is nonsingular and it is estimated by applying a PCA on the empirical variance covariance matrix of ϵ_t. See Alexander (2001), for example.

REFERENCES

Aas, K., C. Czado, A. Frigessi, & H. Bakken (2006) Pair-Copula Constructions of Multiple Dependence. Working paper 487, Munich University.

Aielli, G.P. (2013) Dynamic conditional correlation: On properties and estimation. *Journal of Business & Economic Statistics* 31, 282–299.

Alexander, C. (2001) Orthogonal GARCH. In C. Alexander (ed.), *Mastering Risk*, pp. 21–28. Financial Times-Prentice Hall.

Asai, M., M. McAleer, & J. Yu (2006) Multivariate stochastic volatility: A review. *Econometric Reviews* 25, 145–175.

Bauwens, L., S. Laurent, & J. Rombouts (2006) Multivariate GARCH models: A survey. *Journal of Applied Econometrics* 21, 79–109.

Bedford, T. & R. Cooke (2002) Vines: A new graphical model for dependent random variables. *Annals of Statistics* 30(4), 1031–1068.

Billio, M. & M. Caporin (2006) A Generalized Dynamic Conditional Correlation Model for Portfolio Risk Evaluation. Working paper, Department of Economics, University of Venice.

Bollerslev, T. (1986) Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics* 31, 307–327.

Bollerslev, T. & J.M. Wooldridge (1994) Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric Reviews* 11, 143–172.

Brechmann, E.C. & U. Shepsmeier (2013) Modeling dependence with C- and D-vine copulas: The R Package CDVine. *Journal of Statistical Software* 52, 1–27.

Brechmann, E.C. & H. Joe (2015) Truncation of Vine Copulas Using Fit Indices. Working paper, TUM.

Cambanis, S., S. Huang, & G. Simons (1981) On the theory of elliptically contoured distributions. *Journal of Multivariate Analysis* 11, 368–385.

Cooke, R.M., H. Joe, & K. Aas (2010) Vines arise. In D. Kurowicka & H. Joe (eds.), *Dependence Modeling*: *Handbook on Vine Copulae*, pp. 37–71. World Scientific Publishing Co.

Creal, D., S.J. Koopman, & A. Lucas (2011) A dynamic multivariate heavy-tailed model for time-varying volatilities and correlations. *Journal of Business & Economic Statistics* 29, 552–563.

Daniels, M.J. & M. Pourahmadi (2009) Modeling covariance matrices via partial autocorrelations. *Journal of Multivariate Analysis* 100, 2352–2363.

Diebold, F.X. & R.S. Mariano (1995) Comparing predictive accuracy. *Journal of Business and Economic Statistics* 13, 253–263.

Ding, Z. & R.F. Engle (2001) Large scale conditional covariance matrix modeling, estimation and testing. *Academia Economic Papers* 29, 157–184.

Dissmann, J., E.C. Brechmann, C. Czado, & D. Kurowicka (2013) Selecting and estimating regular vine copulae and application to financial returns. *Computational Statistics and Data Analysis* 59, 52–69.

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

Engle, R.F., V. Ng, & M. Rothschild (1990) Asset pricing with a factor-ARCH covariance structure: Empirical estimates for treasury bills. *Journal of Econometrics* 45, 213–237.

Engle, R.F. & K. Kroner (1995) Multivariate simultaneous GARCH. *Econometric Theory* 11, 122–150.

Engle, R.F. & K. Sheppard (2001) Theoretical and Empirical Properties of Dynamic Conditional Correlation Multivariate GARCH. Working paper No. 8554, National Bureau of Economic Research.

Engle, R.F. & R. Colacito (2006) Testing and valuing dynamic correlations for asset allocation. *Journal of Business and Economic Statistics* 24, 238–253.

Francq, C. & J.M. Zakoïan (2004) Maximum likelihood estimation of pure GARCH and ARMA-GARCH processes. *Bernoulli* 10(4), 605–637.

Gourieroux, C., A. Monfort, & A. Trognon (1984) Pseudo maximum likelihood: Theory. *Econometrica* 52, 681–700.

Hamao, Y., R. Masulis, & V. Ng (1990) Correlations in price changes and volatility across international stock markets. *Review of Financial Studies* 3, 281–307.

Joe, H. (2006) Generating random correlation matrices based on partial correlations. *Journal of Multivariate Analysis* 97, 2177–2189.

Koutmos, G. & G.G. Booth (1995) Asymmetric volatility transmission in international stock. *Journal of International Money and Finance* 14, 747–762.

Kurowicka, D. (2011) Optimal truncation of vines. In D. Kurowicka & H. Joe (eds.), *Dependence Modeling*: *Handbook on Vine Copulae*, pp. 233–248. World Scientific Publishing Co.

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

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

Kurowicka, D. & H. Joe (2010) *Dependence Modelling, Vine Copula Handbook*. World Scientific.

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

Liao, A. & J. Williams (2004) Volatility transmission and changes in stock market interdependence in the European Community. *European Review of Economics and Finance* 3, 203–231.

Patton, A.J. & K. Sheppard (2009) Evaluating volatility and correlation forecasts. In T. Mikosch, J.-P. Kreib, R.A. Davis, & T.G. Andersen (eds.), *Handbook of Financial Time Series*, pp. 801–838. Springer Berlin Heidelberg.

Poignard, B. & J.D. Fermanian (2016a), Dynamic Asset Correlations Based on Vines. Available on SSRN: https://ssrn.com/abstract=2561029.

Poignard, B. & J.-D. Fermanian (2016b) Vine-GARCH Process: Stationarity and Asymptotic Properties. Working paper CREST 2016-03. Available on https://ideas.repec.org/p/crs/wpaper/2016-03.html.

So, M.K.P. & C.Y.T. Yeung (2014) Vine-copula GARCH model with dynamic conditional dependence. *Computational Statistics and Data Analysis* 76, 655–671.

Tse, Y.K. & A.K.C. Tsui (2002) A multivariate GARCH model with time-varying correlations. *Journal of Business and Economic Statistics* 20, 351–362.

Veraverbeke, N., M. Omelka, & I. Gijbels (2011) Estimation of a conditional copula and association measures. *Scandinavian Journal of Statistics* 38, 766–780.

White, H. (1994) *Estimation Inference and Specification Analysis*. Cambridge University Press.

Zhang, W., C.L. Leng, & C.Y. Tang (2015) A joint modelling approach for longitudinal studies. *Journal of the Royal Statistical Society B* 77, 219–238.

APPENDIX A: Some Competing M-GARCH Models

Several (R_t) dynamics have been proposed in the literature. All of them have to cope with the positive definiteness of the correlation matrix and should not depend on too many parameters. The time-varying correlation model of Tse and Tsui (2002) and the DCC model (Engle and Sheppard, 2001) were the first attempts to model dynamic correlations. In this study, we consider the latter as our benchmark.

The DCC model specifies dynamics of the covariance matrix of the de-garched returns u_t directly. In its full form, called "Full DCC", the model belongs to the MARCH family of Ding and Engle (2001) and is specified as

$$Q_t = \left(\iota\iota' - A - B\right) \odot S + A \odot u_{t-1}u'_{t-1} + B \odot Q_{t-1},\ R_t = Q_t^{\star -1/2} Q_t Q_t^{\star -1/2},$$

where $Q_t = \left[q_{ij,t}\right]$ and $Q_t^{\star} = \mathrm{diag}\left(q_{11,t}, q_{22,t}, \ldots, q_{NN,t}\right)$. Above, S, A and B denote $N \times N$ symmetric matrices of unknown parameters and $\odot$ is the usual Hadamard product of two identically sized matrices. Following Ding and Engle (2001), if $\left(\iota\iota' - A - B\right) \odot S$, A and B are positive semidefinite, then the matrix Q_t is positive semidefinite. The significant downside of the full DCC model is its intractability as the (Q_t) process encompasses $3N(N+1)/2$ 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 times the identity matrix. Moreover, the unknown matrix S is most often estimated by the empirical covariance matrix of the standardized returns (the so-called "correlation targeting" technique).

Billio and Caporin (2006) devised 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

$$Q_t = C'SC + A'u_{t-1}u'_{t-1}A + B'Q_{t-1}B,\ R_t = Q_t^{\star -1/2} Q_t Q_t^{\star -1/2},$$

where S, A, B and C are unknown matrices, S being symmetric positive. This model allows for interdependence across groups of assets. The correlation matrices are positive definite 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 corresponds to a model with $3N$ unknown parameters, after correlation targeting.

The set of correlation parameters of the DCC is $\theta_c = (S, A, B)$, whereas for the QFDCC it is $\theta_c = (S, C, A, B)$. In the literature, DCC-GARCH models with correlation 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 procedure produces biased estimates in general and proposed a corrected version of the model called cDCC. Actually, the scalar DCC and cDCC specifications empirically provide very close results. Therefore, in our empirical study, we have considered the scalar DCC and the diagonal QFDCC.

Beside DCC-type models, the GAS family (Creal et al. 2011) provides rather natural alternatives. In such specifications, the asset return dynamics are specified through time-varying vectors of parameters (ϕ_t) of size m, that satisfies the updating equation

$$\phi_t = \omega + As_{t-1} + B\phi_{t-1},$$

where A, B are $m \times m$ matrices of parameters, ω is an unknown m-column vector, and s_t are score-type terms (see Thm. 1 in Creal et al. 2011). Here, in the first step, we set the univariate conditional variances as GARCH(1,1). Conditionally on these first step estimates, we estimate a scalar GAS(1,1) dynamics where $A = a \times Id_m$ and $B = b \times Id_m$, with two positive unknown real numbers a and b.

In our experiments, we have considered two GAS specifications: the hyperspherical GAS and the projected GAS. The hyperspherical GAS corresponds to the model of Creal et al. (2011) stated in their Section 4.2. In this case, ϕ_t is a vector of $m = N(N-1)/2$ angles, that parameterizes the upper-triangular matrices T_t given by the Cholevski decomposition of R_t ($R_t = T_t'T_t$). Therefore, the non-negativeness of R_t is guaranteed by construction. In the projected GAS model, $\phi_t = vech(Q_t)$ so that $m = N(N+1)/2$ and $R_t = Q_t^{*-1} Q_t Q_t^{*-1}$ as previously. Since Q_t may not be positive, we project this matrix on the subset of positive matrices when necessary. This has been simply done by diagonalizing Q_t. When an eigenvalue of Q_t is negative, it is replaced by 0.001. When $N = 2$ and $N = 6$, it has never been necessary to use this projection trick in our simulation experiments. When $N = 10$, Q_t has been replaced by its projected value in 5% of the cases, approximately.

We applied a correlation targeting method for the projected GAS model, a necessary procedure highlighted by Creal et al. (2011), exactly as in the DCC estimation procedure. As for the hyperspherical GAS process, we optimize the second step likelihood function with respect to ω, a and b simultaneously. Here, a targeting procedure is not adapted due to the nature of this GAS specification and we are not aware of a relevant estimate for the unconditional mean of the angles ϕ_t. As in the case of DCC models, one drawback is the growing model complexity with the dimension N as we need to estimate $N(N-1)/2$

parameters among ω. That is why we have estimated both GAS dynamics for $N = 2, 6, 10$ only in Section 4.1.

APPENDIX B: Vine Toolkit

Let $\mathcal{N}$ be a set of n elements $(x_1, \ldots, x_n)$ called nodes. Here, they will correspond to the components of some random vector. An edge e will be an unordered pair of nodes, i.e., $e = (x_i, x_j)$ for some $i, j \in \{1, \ldots, n\}$, $i \neq j$. Consider a set of edges $\mathcal{E}$ associated with $\mathcal{N}$. A path is a sequence of nodes that are connected by edges, i.e., $(x_{i_1}, \ldots, x_{i_p})$ for some indices $i_1, \ldots, i_p$ in $\{1, \ldots, n\}$ and some p, and there is an edge between x_{i_k} and $x_{i_{k+1}}$, $k = 1, \ldots, p-1$. A cycle is a path for which the initial and the final nodes are the same ($i_1 = i_p$ with the previous notations). By definition, $T = (\mathcal{N}, \mathcal{E})$ is a connected tree with nodes $\mathcal{N}$ and edges $\mathcal{E}$ if there is no cycle in the tree and if there is a path between each pair of nodes. Moreover, a vine on n elements is a graph that nests a set of some connected trees $T_1, \ldots, T_{n-1}$, where the edges of tree T_j are the nodes of tree T_{j+1}, $j = 1, \ldots, n-2$. A *regular vine* (R-vine) on n elements 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 $j = 1, \ldots, n-2$. A formal definition is given below. See Kurowicka and Joe (2010) for a survey and additional results.

DEFINITION B.1. (Bedford and Cooke, 2002)
$V(n)$ *is a labeled* regular vine *on n elements if*:

1. $V(n) = (T_1, T_2, \ldots, T_{n-1})$.
2. T_1 *is a connected tree with nodes* $\mathcal{N}_1 = \{1, 2, \ldots, n\}$ *and edges* $\mathcal{E}_1$. *For* $i = 2, \ldots, n-1$, T_i *is a connected tree with nodes* $\mathcal{N}_i = \mathcal{E}_{i-1}$, *and the cardinality of* $\mathcal{N}_i$ *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 consider only regular vines in this article, and the properties we state hereafter are true for such vines implicitly. 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}. The degree of a node is the number of edges linked with it.

Two particular cases of R-vines are important, traditionally. 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 degrees not higher than 2.

Consider an edge e. Starting from some particular node x of $\mathcal{N}_1$ and walking through the successive trees of the vine (from T_1 to T_2, then to T_3, etc.), it is sometimes possible to reach e. In this case, it is said that x is reachable from e. The elements of $\mathcal{N}_1$ reachable from a given edge via this so-called "membership relation" are 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). Thus, we define the conditioning and conditioned sets formally as follows.

DEFINITION B.2. (Cooke, Joe, and Aas, 2010) *For* $e \in \mathcal{E}_i$, $i \leq n-1$, *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* $U_e^\star$.

DEFINITION B.3. (Cooke et al. 2010) *For* $i = 1,\dots,n-1$, *if* $e \in \mathcal{E}_i$, *it connects two elements* j *and* k *in* $\mathcal{N}_i$ *and it can be written* $e = \{j,k\}$. *The* conditioning set *associated with e is* $L_e := U_j^\star \cap U_k^\star$, *and the* conditioned set *associated with e is a pair* $\left\{C_{e,j}, C_{e,k}\right\} := \left\{U_j^\star \setminus L_e, U_k^\star \setminus L_e\right\}$.

Obviously, since the edges of a given tree T_i are the nodes of T_{i+1}, the same concepts of constraint/conditioning/conditioned sets apply to all the nodes in a vine.

LEMMA B.1. (Bedford and Cooke, 2002)
Consider a regular vine on n variables. Then,

1. *the total number of edges is* $n(n-1)/2$;
2. *two different edges have different constraint sets*;
3. *each conditioned set is a doubleton and each pair of variables occurs exactly once as a conditioned set*;
4. *if* $e \in \mathcal{E}_i$, *then* $\#U_e^\star = i+1$, $\#L_e = i-1$;
5. *if two edges have the same conditioning set, then they are the same edge.*

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 $\{a_j, a_k\}$. Both C-, D-, and R-vine and the concepts above can be visualized on Figures 1, 2, and 3.

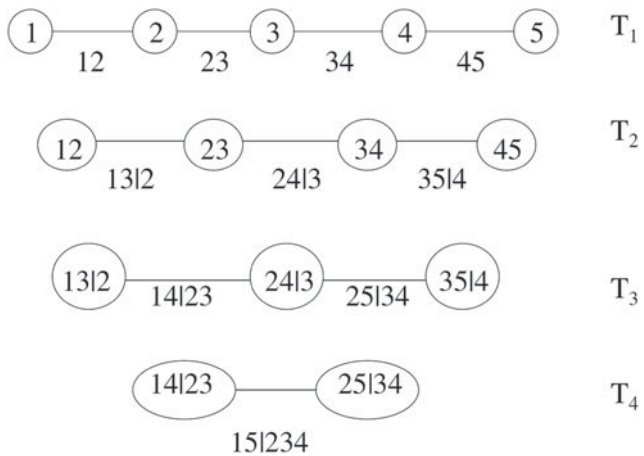

FIGURE 2. Example of a D-vine on five variables. The two nodes $(1,3|2)$ and $(2,4|3)$ in T_3 are connected by the edge $(1,4|2,3)$, whose constraint set is $\{1,2,3,4\}$, conditioned set is $\{1,4\}$ and conditioning set is $\{2,3\}$.

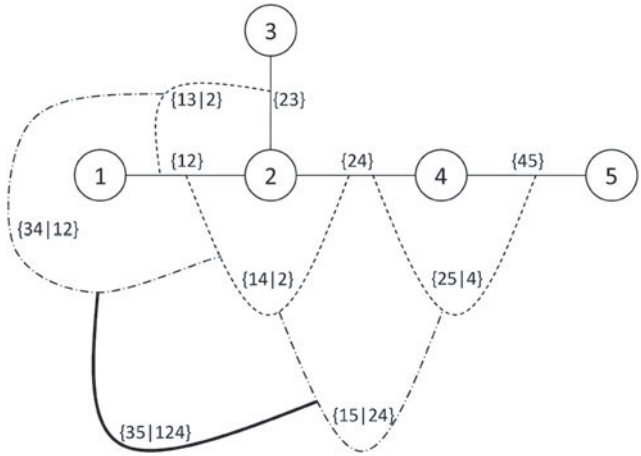

FIGURE 3. Example of an R-vine on five variables. The solid, dotted, dashed-dotted, and black solid lines correspond to the edges of T_1, T_2, T_3, and T_4, respectively.

To illustrate the previous concepts, consider Figure 3, for which $n = 5$. The successive sets of nodes are $\mathcal{N}_1 = \{1, 2, 3, 4, 5\}$, $\mathcal{N}_2 = \{(12), (23), (24), (45)\}$, $\mathcal{N}_3 = \{(13|2), (14|2), (25|4)\}$, $\mathcal{N}_4 = \{(34|12), (15|24)\}$. The edge sets are $\mathcal{E}_1 = \mathcal{N}_2$, $\mathcal{E}_2 = \mathcal{N}_3$,$\mathcal{E}_3 = \mathcal{N}_4$,$\mathcal{E}_4 = \mathcal{N}_5$. For example, the edge (34|12) has connected the nodes (13|2) and (14|2). Its constraint set is $\{1, 2, 3, 4\}$, its conditioning set is $\{1, 2\}$ and its conditioned set is $\{3, 4\}$. In the first tree, the node 2 has degree three, the node 4 has degree two, and all the other nodes have degree one.

To aid intuition, keep in mind that a node represents a random variable, and an edge between two nodes means modelers specify the dependence between these two particular nodes, in general through a copula that has to be defined independently from the vine structure. In the particular case of partial correlation vines (pc-vines, to be short), such copulae are parameterized by partial correlations (Gaussian copula, e.g.). Typically, the goal is to describe the joint law of the n asset returns. For instance, in Figure 1, the five nodes in T_1 may be the asset returns r_i, $i = 1, \ldots, 5$, associated with stock indices. The first tree tells us we will specify the dependencies between r_1 and the other returns r_i, $i > 1$. Here, we select 1 as the core index (the "main factor") in this portfolio. Once we have controlled the T_1-related dependencies, the new nodes in T_2 are conditional asset returns given r_1. We select asset 2 given 1 as the "most relevant" one. The new edges tell us we focus now on conditional copulae between the latter node and the returns r_j given r_1, $j = 2, \ldots, 5$. And we go on with T_3, dealing with the asset returns r_j given r_1 and r_2, $j = 3, 4, 5$, etc. With such a C-vine and a set of convenient bivariate copulae, we obtain the joint law of $(r_1, \ldots, r_5)$ by gathering and multiplying conveniently all the (conditional) copulae we have considered above. This is the simplest way of building vines and joint distributions. Obviously, more complex structures may be relevant too, as in the R-vine of Figure 3. With heterogeneous portfolios, for instance, it would be fruitful to particularize several nodes in T_1. See Aas et al. (2006) for other insights. In terms of model specification, the first chosen trees are crucial because they correspond to our intuitions (our "priors") about the most important linkages among the assets in the portfolio.

Consider a partial correlation vine. From some level on and in practice, it is often possible and useful to assume no dependencies: see the "r-vine free" property in Definition B.5 below.

DEFINITION B.4. *Let a pc-vine* $V(n) = (T_1, T_2, \ldots, T_{n-1})$. *The set of partial correlations associated with this pc-vine is denoted by* $\tilde{C}_{V(n)} := \big(C(T_1), C(T_2), \ldots, C(T_{n-1})\big)$. *Denote by* $R(\tilde{C}_{V(n)})$ *the set of usual correlations that are deduced from* $\tilde{C}_{V(n)}$.

Whatever the values of the partial correlations $\tilde{C}_{V(n)}$ associated with a regular pc-vine $V(n)$, we get a true correlation matrix with the coefficients $R(\tilde{C}_{V(n)})$ (Theorem 1). Since a standardized Gaussian random vector is fully specified by its correlation matrix, we obtain its joint law once we have chosen pc-vine specification. By contrast, for any Gaussian vector, there are many corresponding pc-vine specifications. In a Gaussian world, we recover the interpretation of vines as descriptors of random vector distributions. But more generally, pc-vine specifications can be associated with any random vector, just to describe its correlation matrix (when it exists).

Let us revisit Figure 1 under a partial correlation point of view. To interpret partial correlations, consider linear regressions of some conditioned sets on their conditioning sets. For instance, the node $(1,2)$ and the node $(1,3)$ are connected, and the model will specify the partial correlation $\rho_{12|3}$. This is the correlation between the residuals of the linear regressions of r_2 and r_3 on r_1. Roughly, this measures to what extent r_2 and r_3 are "dependent" given r_1. In practical terms, an econometrician could classify the portfolio components by their (a priori) order of importance. This order may depend on the final phenomenon that is modelled. For instance, if the portfolio payoff depends strongly on emerging markets, it may be relevant to select "Russia" or "Brazil" first instead of "the United States". Intuitively, the latter strategy is intermediate between a factor model where we would regress any asset return on a few prespecified ones, and a PCA where the factors are linear combinations of all returns.

This way of interpreting C-vines has to be revisited with D-vines or even general R-vines. Roughly, D-vines are based on an ordered vision of dependencies across asset returns: any asset is associated with one or two neighbors, with whom correlations are relatively strong. Once they are controlled, the main remaining risk is measured by the correlation with (one or) two other known assets, etc. Such a linear view of the strength of dependencies is probably unrealistic in finance. By contrast, R-vines allow very general and flexible hierarchies and orders among the sequences of partial correlations of interest. Virtually, they allow to integrate any a priori "prior" information, as long as it is consistent with the proximity condition.

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

DEFINITION B.5. *We say that a pc-vine is r-VF* (*VF for vine-Free*) *if*

$$R(C(T_1), C(T_2), \ldots, C(T_{n-1})) = R(C(T_1), C(T_2), \ldots, C(T_{r-1}), C(T'_r), \ldots, C(T'_{n-1})),$$

for any alternative pc-vine $V'(n) := (T_1, T_2, \ldots, T_{r-1}, T'_r, \ldots, T'_{n-1})$, *where the partial correlations associated with the edges of* T'_k, $k \geq r$, *are zero.*

If a pc-vine is r-VF, once the partial correlations are zero above the level r, the correlations are independent of the way this pc-vine has been built from this level. This r-VF property actually holds for any regular pc-vine. This is a consequence of Theorem 2.3 in Brechmann and Joe (2015). They observed that the density of an underlying Gaussian vector is not altered when choosing arbitrary trees $T_{r+1}, \ldots, T_{n-1}$ with associated zero partial correlations.

APPENDIX C: Practical Implementation Details

The computation time is a challenge within a multivariate nonlinear framework. The main advantage of the vine-GARCH model with respect to most competitors is the iterative estimation procedure. This feature allows for optimizing a loss function over 3 parameters—the intercept, the autoregressive, and innovation parameters—for each individual partial correlation process representing a specific node in the vine tree. The estimation can thus be parallelized at each tree level since the usual correlation depends on the corresponding partial correlation of the same tree level and on the partial correlations of the previous trees (see Section 3.3). Each partial correlation process is then estimated independently and the corresponding correlation matrix is positive-definite by construction.

The estimation of an unconstrained vine cannot be performed at once when $N \geq 6$ due to the model complexity because there would be $3N(N-1)/2$ free parameters. Consequently, the iterative procedure is definitely relevant to solve this problem, although the estimation error of the previous estimated partial correlation processes may impact the estimation of the current partial correlation process. Neither this iterative feature nor the associated parallelization can be applied to DCC-type or GAS-type models. Therefore, only simplified form dynamics are empirically feasible in their case, such as scalar versions.

As for the optimization step, we used a Quasi-Newton–Raphson procedure where the parameters are initially specified within "intuitive" subsets. The initial intercept terms ω_{ij} are randomly drawn from a $\mathcal{U}([10^{-4}, 10^{-2}])$ and the initial autoregressive terms ξ_{ij} follow a $\mathcal{U}([0.85, 0.95])$. For the $N-1$ first components of the partial correlation vector—corresponding to the $N-1$ nodes of the first vine tree—, the innovation parameter λ_{ij} is randomly drawn from a $\mathcal{U}([0.05, 0.15])$. Then, from the N-th partial correlation components on, we draw its initial values from a $\mathcal{U}([0.01, 0.05])$. We applied this procedure using the fact that the higher the tree level is, the more we filter the information contained in the conditioning sets. Hence, the partial correlations are expected to progressively become flatter and flatter, and closer to zero, in general.

As we mentioned, the time constraint requires the implementation of the iterative procedure when $N \geq 6$, which is parallelized. Concerning the simulation experiment, when $N = 10$, we estimate 135 parameters. On a PC with four cores and a 2.1GHz processor, the estimation stage lasts approximately 4 hours and 30 minutes when $T = 10,000$. The optimization procedure becomes longer for each new tree level since more computations are required to obtain correlations from partial correlations. As for the application to real portfolios, when $N = 5$ and $T = 3,200$ daily observations, we need to estimate 30 parameters and the iterative estimation procedure lasted 36 minutes.