Série des Documents de Travail

n° 2016-03
Vine-GARCH process: Stationarity and Asymptotic Properties

B.Poignard[1]
J-D.Fermanian[2]

[1] Ensae-Crest & University Paris-Dauphine. E-mail : Benjamin.poignard@ensae.fr

[2] Ensae-Crest. E-mail : Jean-David.Fermanian@ensae.fr

Vine-GARCH process: Stationarity and Asymptotic Properties

Benjamin Poignard* and Jean-David Fermanian†

February 15, 2016

Abstract

We provide conditions for the existence and the uniqueness of strictly stationary solutions of the Vine-GARCH process. The proof is based on Tweedie's (1988) criteria, after rewriting the Vine-GARCH process as a nonlinear Markov chain. Furthermore, we provide asymptotic results of the estimators obtained by the quasi-maximum likelihood method. We prove the weak consistency and asymptotic normality of the quasi-maximum likelihood estimator obtained in a two-step procedure.

Keywords: Asymptotic normality, Consistency, Quasi Maximum Likelihood Estimator, Stationarity, Vine-GARCH.

*Ensae-Crest & University Paris-Dauphine
†Ensae-Crest.

1 Introduction

This paper is the companion one of the working paper CREST "'2014-46, *Dynamic Asset Correlations Based on Vines*, B. Poignard and J.D. Fermanian"'.

The Vine-GARCH process provides an alternative to Dynamic Conditional Correlation (DCC)-type models for specifying the joint dynamics of vectorial stochastic processes. Within the MGARCH framework, the Vine-GARCH specification is a new method for generating dynamics of conditional correlation matrices between asset returns. These correlation matrices are parameterized by a subset of their partial correlations, whose structure are described by an undirected graph called vine. Since such partial correlation processes can be specified separately, our approach provides very flexible and potentially parsimonious multivariate processes. Lewandowski and al. (2009) explained how to deduce a correlation matrix from a partial correlation matrix (or the opposite), through an iterative algorithm. Once the indices of a family of partial correlations is chosen conveniently, a true correlation matrix is generated for any values of these partial correlations. By generating univariate dynamics of partial correlations independently, we obtain sequences of correlation matrices without any normalization stage, contrary to DCC models. The Vine-GARCH model is estimated by a two-step quasi-maximum likelihood procedure.

We prove the existence of stationary solutions, which is the first step towards providing asymptotic results (consistency/asymptotic normality of QML estimates), because law of large numbers (potentially uniform) and some Central Limit Theorems are obtained easily in this case. In the GARCH literature, proving stationarity properties has been fulfilled notably by Bougerol and Picard (1992) for univariate GARCH models, by Ling and McAleer (2003) for multivariate ARMA-GARCH models, by Boussama et al. (2011) for BEKK models, notably. Then we prove the weak consistency of the two-step quasi-maximum likelihood estimator.

After introducing some notations, we specify the Vine-GARCH model. It is rewritten as almost linear Markov chains in Subsection 2.1. The existence of strong and weak stationary solutions is stated in Subsection 2.2. Subsection 2.3 exhibits sufficient conditions to get their uniqueness. Furthermore, consistency and asymptotic normality of the two-step quasi-maximum likelihood estimator are proved respectively in subsection 3.1 and 3.2.

1.1 Notations

Let $A \in \mathcal{M}_{n\times m}(\mathbb{R})$.

- If $n = m$, then $\mathrm{diag}(A) = (a_{ij}\mathbf{1}_{i=j})_{1\le i\le m,1\le j\le m}$ and $\mathrm{Vecd}(A) = (a_{ii})_{1\le i\le m} \in \mathbb{R}^m$.
- If $n = m$ and A symmetric, $\mathrm{Vech}(A) \in \mathbb{R}^q$ with $q = m(m+1)/2$ such that the components are those of A column-wise without redundancy.
- If $n = m$, then $\rho(A)$ is the spectral radius of A, that is the largest of the modulus of the eigenvalues of A. We denote $\lambda_1(A)$ the smallest eigenvalue of A positive definite.
- The Kronecker product is denoted $\otimes$ and $A^{\otimes k} = A \otimes A \otimes \cdots \otimes A$ (k times). The Hadamard product is denoted $\odot$.
- In the following, we consider the submultiplicative norm

  $$\|A\| := \sup\{\frac{\|Ax\|}{\|x\|},\ x \neq 0\},$$

  where $x \in \mathbb{R}^m$ and $\|x\|$ is the Euclidean norm of vector x. For $B \in \mathcal{M}_{m\times n}(\mathbb{R})$, this norm satisfies

  $$\|AB\| \le \|A\|\|B\|, \quad \mathrm{Trace}\,(AB) \le (nm)^{1/2}\|A\|\|B\|.$$

  We define the spectral radius norm for squared non-negative matrices, which is submultiplicative, as

  $$\|A\|_s := \sup\{\sqrt{\lambda} : \lambda \in \mathrm{Spect}\,\big(A'A\big)\}.$$

  We also define the infimum norm of a matrix $A \in \mathcal{M}_{m\times n}(\mathbb{R})$ as

  $$\|A\|_\infty = \max_i \sum_j |A_{ij}|.$$

- For a N dimensional vectorial process $(\epsilon_t)_t$, we denote $\epsilon_t = (\epsilon_{1,t}, \cdots, \epsilon_{N,t})'$ and $\vec{\epsilon}_t := \left(\epsilon_{1,t}^2, \cdots, \epsilon_{N,t}^2\right)'$.
- We denote by $C_b^0(\mathbb{E})$ the space of all continuous and bounded functions $f : \mathbb{E} \to \mathbb{R}$.
- The proofs of consistency and asymptotic normality require some matrix computations, in particular the differentiation of some quantities involving matri-

ces. Recalling some results recorded in Lütkepohl (1996), we have

$$
\begin{aligned}
\frac{\partial x' X x}{\partial X} &= xx',\ X \in \mathcal{M}_{m\times m}(\mathbb{R}), x \in \mathbb{R}^m, \\
\frac{\partial \text{Trace}\,(AX'B)}{\partial X} &= BA,\ X \in \mathcal{M}_{m\times n}(\mathbb{R}), A \in \mathcal{M}_{p\times n}(\mathbb{R}), B \in \mathcal{M}_{m\times p}(\mathbb{R}) \\
\frac{\partial \text{Trace}\left(AX^{-1}B\right)}{\partial X} &= -\left(X^{-1}BAX^{-1}\right)',\ X \in \mathcal{M}_{m\times m}(\mathbb{R}),\ \text{nonsingular}, A, B \in \mathcal{M}_{m\times m}(\mathbb{R}), \\
\frac{\partial \log\left(\det(X)\right)}{\partial X} &= (X')^{-1},\ X \in \mathcal{M}_{m\times m}(\mathbb{R}),\ \text{nonsingular}, \\
\frac{\partial X^{-1}}{\partial x} &= -(X')^{-1}(\partial_x X)X^{-1},\ X \in \mathcal{M}_{m\times m}(\mathbb{R}),\ \text{nonsingular}.
\end{aligned}
$$

1.2 Model Specification

We turn to the Vine-Garch specification. We consider a N-dimensional vectorial stochastic process $(r_t)_{t=1,\cdots,T}$ and denote by θ the vector of the model parameters and decompose the stochastic process $(r_t)_{t=1,\cdots,T}$ as the sum of conditional expected returns and a residual

$$
\begin{aligned}
r_t &= \mu_t(\theta) + \epsilon_t, \\
\epsilon_t &= H_t^{1/2}(\theta)\,\eta_t.
\end{aligned}
\tag{1.1}
$$

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

We focus on the detrended dynamics (ϵ_t). To remove the first moment, we suppose simply that the conditional expected returns are modeled as AR(1), i.e. there exist Φ_0 a $N \times 1$ matrix and Φ_1 a $N \times N$ diagonal matrix s.t. $\mu_t(\theta) = \Phi_0 + \Phi_1 r_{t-1}$. Since we are interested in ϵ_t in this paper, we estimate μ_t by OLS and subtract it from r_t. Now, these estimated residuals will be considered as our observations. The information set is defined by $\mathcal{F}_t = \sigma(r_s, s \leq t) = \sigma(\epsilon_s, s \leq t)$.

The quantity of interest is H_t, which is split between volatility terms contained in D_t and correlation terms in R_t as

$$
H_t = D_t R_t D_t, \tag{1.2}
$$

where $D_t = \text{diag}\left(\sqrt{h_{11,t}}, \cdots, \sqrt{h_{NN,t}}\right)$ is the diagonal matrix of the conditional variances, which is $\mathcal{F}_{t-1}$ measurable. We suppose univariate GARCH dynamics for these conditional variances without cross-effects, such that

$$\text{Vecd}(D_t^2) = V + A.\text{Vecd}(D_{t-1}^2) + B.\vec{\epsilon}_{t-1}, \tag{1.3}$$

where the matrices A and B are diagonal and V is a positive vector of R^N.

The Vine-Garch specification parametrizes the correlation dynamics as

$$\begin{aligned} R_t &= \text{vechof}\left(F_{vine}\left(Pc_t\right)\right), \\ \Psi\left(Pc_t\right) &= \Omega + \Xi\Psi\left(Pc_{t-1}\right) + \Lambda\zeta_{t-1}, \end{aligned} \tag{1.4}$$

where

- vechof$(\cdot)$ denotes the operator "devectorization", that transforms a vector into a symmetric matrix. It is the opposite of the usual operator $vech(\cdot)$.
- Ξ and Λ are $N(N-1)/2 \times N(N-1)/2$ diagonal matrices of unknown parameters, and Ω is an $N(N-1)/2$ unknown vector. Set the vector of parameters $\theta_{cor} = (\Omega, \Xi, \Lambda)$.
- The vector Pc_t is the "partial correlation vector" deduced from a given R-vine structure.
- We apply an analytic transformation Ψ to Pc_t. For the sake of simplicity, Ψ will be known, even if the methodology can be adapted easily to a parametric function Ψ_θ. To fix the ideas, the multivariate Ψ function will be defined as follows:

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

  The function Ψ twists the univariate dynamics to manage the constraints that partial correlations stay between (-1) and 1. Alternatively, Ψ could be chosen among the sigmoïd functions for instance, for which $\psi\left(x\right) = \left(\exp\left(\alpha x\right) - 1\right) / \left(\exp\left(\alpha x\right) + 1\right)$, $\alpha \in \mathbb{R}$.
- The function F_{vine} corresponds to the one-to-one mapping from the vector of partial correlations Pc_t to correlations (in R_t) by using the algorithm of

Lewandowski, Kurowicka and Joe (2009). It is defined as

$$F_{vine} :]-1,1[^{N(N-1)/2} \longrightarrow]-1,1[^{N(N-1)/2},$$
$$F_{vine}\left(\rho_{1,2,t},\cdots,\rho_{N-1,N|L,t}\right) = \left(\rho_{1,2,t},\cdots,\rho_{N-1,N,t}\right)'.$$

- The vector ζ_t consists of a relevant function of the "innovations", to update our partial correlations at time t. More precisely, ζ_t is a measurable and nonlinear transforms of the vector of t-innovations and some quantities that are $\mathcal{F}_{t-1}$ measurable.

The statistical inference is based on the pseudo maximum likelihood procedure, in two steps, which is also called quasi maximum likelihood estimator (QMLE). We observe a T-path $(\epsilon_t)_{t=1,\dots,T}$ of the random vector ϵ. Such a process corresponds to a realization drawn following the unique, strict-sense stationary and nonanticipative solution (ϵ_t) of (1.1). To avoid any confusion, we denote by $D_t(\theta)$, $R_t(\theta)$ and $H_t(\theta)$ the (diagonal) t-matrix of conditional volatilities, the matrix of conditional correlations and the conditional variance-covariance matrix respectively, when they are generated by our model equations, and assuming θ is the underlying parameter.

We estimate our model (1.1) by a Gaussian QMLE, assuming the unknown true parameter θ_0 belongs to some compact set Θ. We apply a two-step estimation method, that is usual in this stream of the literature. To do so, we work as if $(\eta_t)_t$ were a Gaussian white noise. Therefore, the likelihood function can be split into two parts: the variance part on one side, and the correlation part on the other side.

We denote $\theta = (\theta_v, \theta_c)$, such that $\theta_v := (\theta_i, i = 1, \cdots, 3N)$ corresponds to the volatility parameters, and $\theta_c := (\theta_i, i = 3N+1, \cdots, 3N + 3N(N-1)/2)$ to the correlation parameters. The variance part of the log-likelihood function is the sum of log-likelihood functions of N univariate GARCH(1,1) models that can be estimated independently:

$$\begin{cases} \hat{\theta}_{T,v} = \underset{\theta_v \in \Theta_v}{\arg\min}\, QL_{1,T}\left(\theta_v;\epsilon\right), \\ QL_{1,T}\left(\theta_v;\epsilon\right) = \frac{1}{T}\sum_{t=1}^{T} l_{1,t}\left(\epsilon_t;\theta_v\right) := \frac{1}{T}\sum_{t=1}^{T}\sum_{i=1}^{N}\left[\log\left(h_{i,t}\right) + \frac{\epsilon_{i,t}^2}{h_{i,t}}\right], \end{cases} \tag{1.5}$$

where the sequences of variances $(h_{i,t})_{t\geq 1}$ are generated under the assumed parameter θ_v. In other words, $D_t(\theta_v) = \operatorname{diag}\left(h_{1,t}^{1/2},\cdots,h_{N,t}^{1/2}\right)$. Above, Θ_v is the projection of the parameter set Θ on the sub-space of the variance-related components. Given $\hat{\theta}_{T,v}$, a consistent (but inefficient) estimator of $\theta_{0,v}$, an estimator of $\theta_{0,c}$ can be obtained as

$$\left\{\begin{array}{lcl}\hat{\theta}_{T,c} & = & \underset{\theta_c \text{ s.t. } (\hat{\theta}_{T,v},\theta_c)\in\Theta}{\arg\min} \; QL_{2,T}\left(\hat{\theta}_{T,v},\theta_c;\epsilon\right), \\ QL_{2,T}\left(\hat{\theta}_{T,v},\theta_c;\epsilon\right) & = & \frac{1}{T}\sum_{t=1}^{T} l_{2,t}(\epsilon_t;\hat{\theta}_{T,v},\theta_c) := \frac{1}{T}\sum_{t=1}^{T}\left[\log\left(|R_t(\hat{\theta}_{T,v},\theta_c)|\right) + \hat{u}_t' R_t^{-1}(\hat{\theta}_{T,v},\theta_c)\hat{u}_t\right],\end{array}\right. \tag{1.6}$$

where $\hat{u}_t = D_t^{-1}(\hat{\theta}_{T,v})\epsilon_t$.

Both previous criteria are C^∞ and the minimization problem comes from the negative sign in the pseudo log-likelihood, which can be ruled out. The orthogonal conditions are

$$\left\{\begin{array}{lcl}\Delta_T(\hat{\theta}_{T,v}) & = & \frac{1}{T}\sum_{t=1}^{T}\delta_t(\hat{\theta}_{T,v}) = 0, \quad \text{with } \delta_t(\theta_{T,v}) := \nabla_{\theta_v} l_{1,t}(\epsilon_t;\theta_v), \\ \Psi_T(\hat{\theta}_{T,v},\hat{\theta}_{T,c}) & = & \frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta}_{T,v},\hat{\theta}_{T,c}) = 0, \text{ with } \psi_t(\theta) := \nabla_{\theta_c} l_{2,t}(\epsilon_t;\theta).\end{array}\right. \tag{1.7}$$

We denote by $\delta_t^{(i)}(\theta_v)$ (resp. $\psi_t^{(i)}(\theta_v,\theta_c)$) the i-th component of $\nabla_{\theta_v} l_{1,t}(\theta_v)$ (resp. $\nabla_{\theta_c} l_{2,t}(\theta_v,\theta_c)$).

An issue is the choice of the initial values to generate (1.1). Indeed, the marginal variance processes and the correlation dynamics need to be initialized at time $t=1$. To do so, we propose to initialized them by their sample counterparts:

$$\forall i = 1,\cdots,N, \; \tilde{h}_{i,1} = \frac{1}{T-1}\sum_{t=1}^{T}\epsilon_{i,t}^2, \; \tilde{D}_1 = \text{diag}(\tilde{h}_{i,1}^{1/2}) \text{ and } \tilde{R}_1 = \frac{1}{T}\sum_{t=1}^{T}\epsilon_1\tilde{D}_1^{-2}\epsilon_1'. \tag{1.8}$$

A volatility process $(\tilde{D}_t)_{t>1} = (\text{diag}(\tilde{h}_{i,t}^{1/2}))_{t>1}$ and a correlation process $(\tilde{R}_t)_{t>1}$ are generated starting from these initial values. Hence, besides the "theoretical" quantities $QL_{1,T}(\theta_v;\epsilon)$ and $QL_{2,T}(\theta_v,\theta_c;\epsilon)$, we denote by $\widetilde{QL}_{1,T}(\theta_v;\epsilon)$ and $\widetilde{QL}_{2,T}(\theta_v,\theta_c;\epsilon)$ the log-likelihoods generated from some fixed initial values (as those proposed above). The same holds for $\tilde{\Delta}_T(\theta_v)$ and $\tilde{\Psi}_T(\theta_v,\theta_c)$, etc.

2 Stationarity

In this section, we specify the Data Generating Process (DGP) differently from the specification given in (1.1). A significant quantity is the vector of standardized residuals, defined as $u_t = D_t^{-1}\epsilon_t$. We straightforwardly have $\mathbb{E}_{t-1}[u_t] = 0$ and $\mathbb{E}_{t-1}[u_t u_t'] = R_t$. This implies that u_t can be specified as $u_t = R_t^{1/2}\eta_t^*$, such that η_t^* is a centered random vector with $\mathbb{E}_{t-1}[\eta_t^*\eta_t^{*\prime}] = I_N$. Therefore, the "true" DGP will be the stationary process (η_t^*). The two "innovations" (η_t) and (η_t^*) are related to each other by the relation

$$H_t^{1/2}\eta_t = D_t R_t^{1/2}\eta_t^*.$$

Note that, if $\mathbb{E}_{t-1}[\eta_t^*] = 0$ and $\mathbb{E}_{t-1}[\eta_t^*\eta_t^{*\prime}] = I_N$, then $\mathbb{E}_{t-1}[\eta_t] = 0$ and $\mathbb{E}_{t-1}[\eta_t\eta_t'] = I_N$, and the opposite.

2.1 Vine-Garch as Markov Chains

The Vine-Garch specification can be written as a Markov chain, a representation that is relevant for studying stationary solutions. To do so, we define

$$X_t := \left(\vec{\epsilon}_t, \mathrm{Vecd}(D_t^2), \Psi\left(Pc_t\right)\right)', \tag{2.1}$$

such that, for all $t > 0$, $(X_t)_t$ satisfies

$$X_t = T_t X_{t-1} + \nu_t. \tag{2.2}$$

This means $(X_t)_t$ follows an autoregressive form of order 1 with stochastic T_t. Let us focus on the first component of X_t. Setting $\vec{u}_t := (u_{1,t}^2, \ldots, u_{N,t}^2)$, we have

$$D_t^2\vec{u}_t = \vec{u}_t \odot \mathrm{Vecd}(D_t^2) = \vec{\epsilon}_t = \vec{u}_t \odot V + \vec{u}_t \odot A.\mathrm{Vecd}(D_{t-1}^2) + \vec{u}_t \odot B.\vec{\epsilon}_{t-1}. \tag{2.3}$$

Using the dynamics of $\mathrm{Vecd}(D_t^2)$ and $\Psi\left(Pc_t\right)$, the matrix T_t satisfies

$$T_t = \begin{pmatrix} \vec{u}_t \odot B & \vec{u}_t \odot A & 0 \\ B & A & 0 \\ 0 & 0 & \Xi \end{pmatrix}, \tag{2.4}$$

and the vector of innovation ν_t is defined as

$$\nu_t = \begin{pmatrix} \vec{u}_t \odot V \\ V \\ \Omega + \Lambda\zeta_{t-1} \end{pmatrix}. \tag{2.5}$$

Note that $\zeta_t = \zeta\left(\chi_t, \eta_t\right)$ where $\chi_t = (Pc_t, D_t)$.

Assumption 1. *The vectorial process $(\eta_t^*)_{t\in\mathbb{Z}}$ satisfies the Markov property with respect to $\mathcal{F}$, i.e*

$$\forall t \in \mathbb{Z},\ \mathbb{E}\left[\eta_t^*|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\eta_t^*|X_{t-1}\right].$$

Besides, $\mathbb{E}_{t-1}\left[\eta_t^*\right] = 0$ *and* $\mathbb{E}_{t-1}\left[\eta_t^*\eta_t^{*\prime}\right] = I_N$.

As a consequence (and equivalently, in fact), the same property is fulfilled with the other "innovations" $(\eta_t)_{t\in\mathbb{Z}}$: the process $(\eta_t)_{t\in\mathbb{Z}}$ satisfies the Markov property with respect to $\mathcal{F}$, i.e

$$\forall t \in \mathbb{Z},\ \mathbb{E}\left[\eta_t|\mathcal{F}_{t-1}\right] = \mathbb{E}\left[\eta_t|X_{t-1}\right].$$

Moreover, $\mathbb{E}_{t-1}\left[\eta_t\right] = 0$ and $\mathbb{E}_{t-1}\left[\eta_t\eta_t'\right] = I_N$.

Proposition 2.1. *Under Assumption 1,* $(X_t)_t$ *is a Markov Chain of order one.*

Proof. Note that $u_t = D_t^{-1}H_t^{1/2}\eta_t$, where H_t is a deterministic function of X_{t-1}. Since η_t satisfies the Markov property with respect to $\mathcal{F}$, then $u_t|\mathcal{F}_{t-1} \overset{d}{=} u_t|X_{t-1}$. Furthermore, X_t can be rewritten as follows: there exists constant matrices Γ_1 and Γ_2 such that

$$X_t = (\Gamma_1.\xi_t) \odot T_0 X_{t-1} + (\Gamma_2.\vec{\chi}_t) \odot \nu_0, \tag{2.6}$$

where T_0 (resp. ν_0) is the T_t (resp. ν_t) matrix when $u_t = 1$, $\xi_t := (\vec{u}_t, 1)'$ and $\vec{\chi}_t := (\vec{u}_t, 1, \zeta(\chi_{t-1}, \eta_{t-1}))'$. Then X_t is a measurable function of $(\eta_t, X_{t-1}, \eta_{t-1})$, where η_t satisfies the Markov property by (1). Consequently, $(X_t)_t$ is Markovian.

□

2.2 Existence of stationary Vine-Garch solutions

The recurrence equation (2.2) is stochastic through T_t and ν_t, i.e. through the innovations η_t (or η_t^*) and the $\mathcal{F}_{t-1}$-measurable matrix R_t. A consequence of this parametrization is that T_t depends on subcomponents of X_t. Hence, we can not extract an expression such as $X_t = f(\eta_t, \eta_{t-1}, \cdots)$ nor $X_t = f\left(\eta_t^*, \eta_{t-1}^*, \cdots\right)$, for some explicit function $f(.)$. This comes from the nonlinear relationship between T_t and the past innovations (before and including t). Classical techniques such as Lyapunov exponent are not adapted in our framework.

The existence of stationary solutions -but not a unique solution- for the vine-GARCH model can be proved using the criterion of Tweedie (1988). Tweedie provides the existence of an invariant probability measure for the Markov chain defined in (2.2). Ling and McAleer used this criterion to establish the stationarity of vector ARMA-GARCH models.

The stationarity of the $(\vec{\epsilon}_t)_t$ process requires the control of T_t, which should avoid non-explosive patterns. The matrix T_t is a function of $(\vec{u}_t)_t$, which are dependent variables. Furthermore, the conditional law of $\vec{u}_t$ is a function of H_t and D_t, which in turn is a function of X_{t-1}. This is the reason we need the next hypothesis.

Assumption 2. *For some $p \geq 1$, $\|T^\star\|_s < \infty$, where*

$$T^\star := \sup_{\mathbf{x}\in\mathbb{R}^d} \mathbb{E}\left[|T_t^{\otimes p}||X_{t-1} = \mathbf{x}\right]. \tag{2.7}$$

Assumption 3. *Denoting by λ the Lebesgue measure, the conditional kernel of η_t^* given $X_{t-1} = \mathbf{x}$ is defined as*

$$d\mathbb{P}_{\eta_t^*}^{X_{t-1}=\mathbf{x}}(u) = f_{\eta_t^*}(u|\mathbf{x})d\lambda(u). \tag{2.8}$$

Furthermore, for all $u \in \mathbb{R}^m$, the mapping $\mathbf{x} \to f_{\eta_t^*}(u|\mathbf{x})$ is continuous and there exists an integrable function g such that, for all $u \in \mathbb{R}^m$,

$$\sup_t \sup_{\mathbf{x}\in\mathbb{R}^d} f_{\eta_t^*}(u|\mathbf{x}) \leq g(u). \tag{2.9}$$

Moreover, $\forall t$, $\mathbb{E}\left[\|\eta_t^*\|^{2p}|X_{t-1} = \mathbf{x}\right] \leq \psi(\|\mathbf{x}\|)$ satisfying $\forall \alpha > 0$, $\lim_{v\to\infty} \frac{\psi(v)}{v^\alpha} = 0$.

Assumption 4. *There exists a positive real number a such that, for almost every trajectory and every $\theta \in \Theta$, the partial correlations of our chosen vine (i.e. the components of the vectors $Pc_t(\theta)$) belong to the fixed interval $[-1 + a, 1 - a]$.*

In particular, the latter assumption implies that, for every $\theta \in \Theta$, the determinant of almost every correlation matrices $R_t(\theta)$ are strictly larger than $a^{N(N-1)} > 0$ (apply Kurowicka and Cooke, 2006, Theorem 3.2), and that the norm of $R_t^{-1}(\theta)$ is bounded from above a.e. [1]. Moreover, the function F_{vine} that maps partial correlations to usual correlations has a bounded derivative, when applied to the trajectories $(Pc_t(\theta))$ generated by the model.

Theorem 2.2. *Under Assumptions 1-4 the process (ϵ_t, D_t, R_t) as defined in equations (1.2), (1.3), and (1.4) possesses a strictly stationary solution such that $(\epsilon_t, D_t, R_t) \in \mathcal{F}_t$, the sigma field induced by the observations. Furthermore, the solution (ϵ_t) is second-order stationary and, when the innovations η_t^* are Gaussian given $\mathcal{F}_{t-1}$, then $\mathbb{E}\left[\|\epsilon_t\|^{2p}\right] < \infty$.*

The key result for the existence of an invariant probability measure for Markov chains is the criterion of Tweedie (1988). When using this approach, the irreducibility of (X_t) is not required to obtain stationarity.

[1]Indeed, $\|R_t^{-1}\|_s \leq \lambda_{min}(R_t)^{-N} \leq a^{N^2(N-1)}$.

Let $(X_t)_{t\in\mathbb{Z}}$ be a homogeneous Markov chain with a measurable state space $(E,\mathcal{E})$, such that its transition probability is $P(\mathbf{x},B)=\mathbb{P}\left(X_t\in B|X_{t-1}=\mathbf{x}\right)$, where $\mathbf{x}\in E$ and $B\in\mathcal{E}$. Theorem 2 of Tweedie (1988) states the following:

Lemma 2.3. *Suppose $(E,\mathcal{E})$ is a locally compact separable state space and $(X_t)_{t\in\mathbb{Z}}$ is a Feller chain, that is for $h\in C_b^0(E)$, then $E\left[h(X_t)|X_{t-1}=\mathbf{x}\right]$ is also $C_b^0(E)$.*

1. *If for some compact set $B\in\mathcal{E}$, there exists a non negative mapping $g(.)$ and $\epsilon>0$ such that*

$$\int_{B^c} P(\mathbf{x},y)g(y)d\lambda(y)\leq g(\mathbf{x})-\epsilon,\ \mathbf{x}\in B^c, \tag{2.10}$$

then there exists a σ-finite invariant measure μ for P such that $0<\mu(B)<\infty$.

2. *Furthermore, if*

$$\int_B\left(\int_{B^c} P(\mathbf{x},y)g(y)d\lambda(y)\right)d\mu(\mathbf{x})<\infty, \tag{2.11}$$

then μ is finite and hence $\pi=\mu/\mu(E)$ is an invariant probability measure.

3. *Furthermore, if*

$$\int_{B^c} P(\mathbf{x},y)g(y)d\lambda(y)\leq g(\mathbf{x})-f(\mathbf{x}),\ \mathbf{x}\in B^c, \tag{2.12}$$

then μ admits a finite f-moment, i.e. $\mathbb{E}_\mu\left[f(X_t)\right]<\infty$.

The next Lemma is a specific version of Lemma A.2 in Ling and McAleer (2003). Its proof is omitted.

Lemma 2.4. *For a given squared matrix T, if $\rho(|T|)<1$, then there exists a positive vector M such that $(Id-|T|)'M>0$.*

Proof. We first show that $(X_t)_{t\in\mathbb{Z}}$ is a Feller process. Let $h\in C_b^0(\mathbb{R}^d)$. We have

$$\begin{aligned}\mathbb{E}\left[h(X_t)|X_{t-1}=\mathbf{x}\right] &= \mathbb{E}\left[h(T_t\mathbf{x}+\nu_t)|X_{t-1}=\mathbf{x}\right]\\ &= \mathbb{E}\left[h(\phi_1(u_t)\mathbf{x}+\phi_2(u_t,\eta^*_{t-1}))|X_{t-1}=\mathbf{x}\right],\end{aligned} \tag{2.13}$$

for continuous transforms ϕ_1 and ϕ_2. By construction, $u_t=D_t^{-1}H_t^{1/2}\eta_t=R_t^{1/2}\eta_t^*$, where $R_t^{1/2}$ is a continuous mapping of X_{t-1}. Consequently, we obtain

$$\begin{aligned}\mathbb{E}\left[h(X_t)|X_{t-1}=\mathbf{x}\right] &= \mathbb{E}\left[h\circ\tilde{\phi}(\mathbf{x},\eta_t^*)|X_{t-1}=\mathbf{x}\right]\\ &= \int h\circ\tilde{\phi}(\mathbf{x},u)d\mathbb{P}_{\eta_t^*}^{X_{t-1}=\mathbf{x}}(u)\\ &= \int h\circ\tilde{\phi}(\mathbf{x},u)f_{\eta_t^*}(u|\mathbf{x})d\lambda(u),\end{aligned} \tag{2.14}$$

for some continuous transform $\tilde{\phi}$. Now, let $(\mathbf{x}_n)_n$ be a sequence such that $\mathbf{x}_n \underset{n\to\infty}{\longrightarrow} \mathbf{x}$. As $h(.)$ is bounded and $\forall u$, $(h\circ\tilde{\phi}(\mathbf{x}_n,u))_n$ is convergent, then $\lim_n \mathbb{E}\left[h(X_t)|X_{t-1}=\mathbf{x}_n\right] = \mathbb{E}\left[h(X_t)|X_{t-1}=\mathbf{x}\right]$ by the Lebesgue dominated convergence theorem under (3). In other words, $\mathbf{x}\to\mathbb{E}\left[h(X_t)|X_{t-1}=\mathbf{x}\right]$ is continuous.

Second, we exhibit an explicit functional $g(.)$ to apply the Tweedie's criteria. To do so, take $g(\mathbf{x}) = 1 + |\mathbf{x}^{\otimes p}|' M$, for any vector M, which will be explicit later. We have, for $p\geq 1$,

$$\mathbb{E}\left[g(X_t)|X_{t-1}=\mathbf{x}\right] = 1 + \mathbb{E}\left[|(T_t\mathbf{x}+\nu_t)^{\otimes p}|'|X_{t-1}=\mathbf{x}\right] M.$$

By some property of the Kronecker product and algebraic manipulations, let us rewrite $(T_t\mathbf{x}+\nu_t)^{\otimes p} = (T_t\mathbf{x})^{\otimes p} + \mathcal{B}(\mathbf{x}) = T_t^{\otimes p}\mathbf{x}^{\otimes p} + \mathcal{B}(\mathbf{x})$. We deduce that

$$\mathbb{E}\left[g(X_t)|X_{t-1}=\mathbf{x}\right] \leq 1 + \left(\mathbb{E}\left[|T_t^{\otimes p}\mathbf{x}^{\otimes p}|'|X_{t-1}=\mathbf{x}\right] + \mathbb{E}\left[\|\mathcal{B}(\mathbf{x})\||X_{t-1}=\mathbf{x}\right]\right) M. \tag{2.15}$$

We focus on the first expectation in (2.15). As T_t is a function of u_t, its conditional distribution depends on R_t. Hence T_t is a function of X_{t-1}. Then, we obtain

$$\begin{aligned}
\mathbb{E}\left[|(T_t\mathbf{x})^{\otimes p}|'|X_{t-1}=\mathbf{x}\right] M &\leq |\mathbf{x}^{\otimes p}|'\mathbb{E}\left[|T_t^{\otimes p}|'|X_{t-1}=\mathbf{x}\right] M \\
&\leq |\mathbf{x}^{\otimes p}|'\left(\sup_{\mathbf{x}\in\mathbb{R}^d}\mathbb{E}\left[|T_t^{\otimes p}|'|X_{t-1}=\mathbf{x}\right]\right) M \\
&\leq |\mathbf{x}^{\otimes p}|'(T^\star)' M.
\end{aligned}$$

As for the second expectation in (2.15), by taking any multiplicative norm $\|.\|$, we have

$$\mathbb{E}\left[\|\mathcal{B}(\mathbf{x})\||X_{t-1}=\mathbf{x}\right] \leq K\mathbb{E}\left[\|\nu_t\|\|(T_t\mathbf{x})^{\otimes(p-1)}\| + \|\nu_t\|^2\|(T_t\mathbf{x})^{\otimes(p-2)}\| + \cdots + \|\nu_t\|^p|X_{t-1}=\mathbf{x}\right], \tag{2.16}$$

where K is a non-negative constant. In (2.16), we need to upper bound quantities of the type $\mathbb{E}\left[\|\nu_t\|^m\|T_t\|^n|X_{t-1}=\mathbf{x}\right]$, i.e. terms as $\mathbb{E}\left[(\|\zeta_{t-1}\|+\|\vec{u}_t\|)^m\|\vec{u}_t\|^n|X_{t-1}=\mathbf{x}\right]$ when $m+n\leq p$. First, we consider $\mathbb{E}\left[\|\vec{u}_t\|^{m+n}|X_{t-1}=\mathbf{x}\right]$. Recall that $u_t = R_t^{1/2}\eta_t^*$. Taking the spectral norm of $R_t^{1/2}$, we obtain a.s.

$$\|R_t^{1/2}\| = \rho\left(R_t^{1/2}R_t^{1/2\,\prime}\right)^{1/2} = \sqrt{\text{Trace}\left(D_t^{-1}H_tD_t^{-1}\right)} \leq \sqrt{N}.$$

Using the previous inequality and Assumption 3, we have

$$\mathbb{E}\left[\|\vec{u}_t\|^{m+n}|X_{t-1}=\mathbf{x}\right] \leq \mathbb{E}\left[\|R_t^{1/2}\|^{2(m+n)}\|\vec{\eta}_t^*\|^{m+n}|X_{t-1}=\mathbf{x}\right] \leq N^{n+m}\mathbb{E}\left[\|\vec{\eta}_t^*\|^{m+n}|X_{t-1}=\mathbf{x}\right]. \tag{2.17}$$

By assumption, $\mathbb{E}\left[\|\eta_t^*\|^{2p}|X_{t-1}=\mathbf{x}\right] \leq \psi(\|\mathbf{x}\|)$. Then, we obtain

$$\mathbb{E}\left[\|\vec{u}_t\|^{m+n}|X_{t-1}=\mathbf{x}\right] \leq \alpha_{m,n}\psi\left(\|\mathbf{x}\|\right)^{(m+n)/p},$$

for some constant $\alpha_{m,n}$.

Another product element we shall bound is $\mathbb{E}\left[\|\zeta\left(\chi_{t-1},\eta_{t-1}\right)\|^m\|\vec{u}_t\|^n|X_{t-1}=\mathbf{x}\right]$. To do so, we take $n+m=p$, where $m\geq 1$. Using the conditional Hölder inequality, we obtain

$$\mathbb{E}\left[\|\zeta\left(\chi_{t-1},\eta_{t-1}\right)\|^m\|\vec{u}_t\|^n|X_{t-1}=\mathbf{x}\right] \leq \mathbb{E}\left[\|\zeta\left(\chi_{t-1},\eta_{t-1}\right)\|^p|X_{t-1}=\mathbf{x}\right]^{m/p}\mathbb{E}\left[\|\vec{u}_t\|^p|X_{t-1}=\mathbf{x}\right]^{n/p}. \tag{2.18}$$

In (2.18), $\mathbb{E}\left[\|\vec{u}_t\|^p|X_{t-1}=\mathbf{x}\right]^{n/p}$ can be straightforwardly upper bounded using (2.17). We now focus on the conditional expectation of $\|\zeta\left(\chi_{t-1},\eta_{t-1}\right)\|^p$. Denoting $\tilde{\upsilon}_{k|L,t}=\epsilon_{k,t}-\mathbb{E}_{t-1}\left[\epsilon_{k,t}|\epsilon_{L,t}\right]$, we have

$$\mathbb{E}\left[\|\zeta\left(\chi_{t-1},\eta_{t-1}\right)\|^p|X_{t-1}=\mathbf{x}\right] \leq \sup_{(i,j|L)\in E}\mathbb{E}\left[|\frac{\tilde{\upsilon}_{i|L,t-1}\tilde{\upsilon}_{j|L,t-1}}{\sqrt{h_{i|L,t-1}}\sqrt{h_{j|L,t-1}}}|^p|X_{t-1}=\mathbf{x}\right]. \tag{2.19}$$

For $p=1$, we apply the Cauchy-Schwartz inequality to (2.19) as

$$\mathbb{E}\left[|\frac{\tilde{\upsilon}_{i|L,t-1}\tilde{\upsilon}_{j|L,t-1}}{\sqrt{h_{i|L,t-1}}\sqrt{h_{j|L,t-1}}}||X_{t-1}=\mathbf{x}\right] \leq \mathbb{E}\left[\frac{\tilde{\upsilon}^2_{i|L,t-1}}{h_{i|L,t-1}}|X_{t-1}=\mathbf{x}\right]^{1/2}\mathbb{E}\left[\frac{\tilde{\upsilon}^2_{j|L,t-1}}{h_{j|L,t-1}}|X_{t-1}=\mathbf{x}\right]^{1/2} = 1.$$

In this case, we obtain

$$\mathbb{E}\left[\|\mathcal{B}(\mathbf{x})\||X_{t-1}=\mathbf{x}\right] = \alpha_1\mathbb{E}\left[\|\zeta_{t-1}\|+\|u_t\||X_{t-1}=\mathbf{x}\right] \leq \alpha_2\psi\left(\|\mathbf{x}\|\right)+\alpha_3, \tag{2.20}$$

for some constants α_k, $k=1,2,3$. Consequently for $p=1$, we deduce that (2.15) can be upper bounded as

$$\begin{aligned}\mathbb{E}\left[g(X_t)|X_{t-1}=\mathbf{x}\right] &\leq 1+\left(\mathbb{E}\left[|T_t\mathbf{x}|'|X_{t-1}=\mathbf{x}\right]+\mathbb{E}\left[\|\mathcal{B}(\mathbf{x})\||X_{t-1}=\mathbf{x}\right]\right)M \\ &\leq 1+|\mathbf{x}|'(T^\star)'M+O\left(\|\mathbf{x}\|^a\right),\end{aligned}$$

for any $a>0$. Let us now try to extend this result for $p>1$. The quantity given in

(2.19) is a product of $\tilde{\upsilon}_{k|L,t-1}$ components, which can be decomposed as

$$\begin{aligned} \tilde{\upsilon}_{i|L,t-1} &= e_i' H_{t-1}^{1/2}(\theta)\{\eta_{t-1} - \mathbb{E}_{t-2}\left[\eta_{t-1}|\epsilon_{L,t-1}, X_{t-1} = \mathbf{x}\right]\} \\ &= e_i' D_{t-1} R_{t-1}^{1/2}\{\eta_{t-1}^* - \mathbb{E}_{t-2}\left[\eta_{t-1}^*|\epsilon_{L,t-1}, X_{t-1} = \mathbf{x}\right]\} \end{aligned} \tag{2.21}$$

Assuming all denominators are bounded from below a.s., this implies that (2.19) can be upper bounded as

$$\begin{aligned} \sup_{(i,j|L)\in E} \mathbb{E}\left[|\frac{\tilde{\upsilon}_{i|L,t-1}\tilde{\upsilon}_{j|L,t-1}}{\sqrt{h_{i|L,t-1}}\sqrt{h_{j|L,t-1}}}|^p | X_{t-1} = \mathbf{x} \right] &\leq Cst.\mathbb{E}\left[\|D_{t-1}\|^{2p}\|R_{t-1}\|^p\|\eta_{t-1}^*\|^{2p}|X_{t-1} = \mathbf{x}\right] \\ &\leq Cst.\mathbb{E}\left[\|\mathbf{x}\|^p\|\eta_{t-1}^*\|^{2p}|X_{t-1} = \mathbf{x}\right] \\ &\leq Cst.\|\mathbf{x}\|^p\psi(\|\mathbf{x}\|). \end{aligned}$$

This upper bound is not of order $O(\|\mathbf{x}\|^k)$, for $k \leq p-1$. We rely on the Gaussian distribution hypothesis to circumvent this obstacle.

Now, the vectors η_t^* (or η_t, equivalently) is supposed to be gaussian, conditional to the past. By the Cauchy-Schwartz inequality, we have

$$\mathbb{E}\left[|\frac{\tilde{\upsilon}_{i|L,t-1}\tilde{\upsilon}_{j|L,t-1}}{\sqrt{h_{i|L,t-1}}\sqrt{h_{j|L,t-1}}}|^p | X_{t-1} = \mathbf{x} \right] \leq \mathbb{E}\left[\frac{\tilde{\upsilon}_{i|L,t-1}^{2p}}{h_{i|L,t-1}^p} | X_{t-1} = \mathbf{x} \right]^{1/2} \mathbb{E}\left[\frac{\tilde{\upsilon}_{j|L,t-1}^{2p}}{h_{j|L,t-1}^p} | X_{t-1} = \mathbf{x} \right]^{1/2}.$$

Since any $\tilde{\upsilon}_{i|L,t-1}/\sqrt{h_{i|L,t-1}}$ is a Gaussian random variable $\mathcal{N}(0,1)$, given X_{t-1}, the r.h.s. of the latter inequality is uniformly bounded wrt i, j, L and $\mathbf{x}$. We deduce that (2.19) can be upper bounded as

$$\mathbb{E}\left[\|\zeta\left(\chi_{t-1}, \eta_{t-1}\right)\|^p | X_{t-1} = \mathbf{x}\right] = O(1), \tag{2.22}$$

for all $\mathbf{x}$.

This result is proved using $\forall t \geq 1, \sigma_{k|L,t}^2(\mathbf{x}) > 0$ a.s.. We need to prove that this holds almost surely for any $\mathbf{x} \in B^c$. That means we need to control for the variance and correlation dynamics when $\mathbf{x}$ can take very large values. By contradiction, suppose $\forall k \notin L$

$$\sigma_{k|L,t}^2(\mathbf{x}) = \mathbb{E}\left[\left(\epsilon_{k,t} - \mathbb{E}\left[\epsilon_{k,t}|\epsilon_{L,t}\right]\right)^2 | X_{t-1} = \mathbf{x}\right] = 0 \Rightarrow \epsilon_{k,t} = \mathbb{E}\left[\epsilon_{k,t}|\epsilon_{L,t}, X_{t-1} = \mathbf{x}\right] \text{ a.s.} \tag{2.23}$$

Using the decomposition $\epsilon_t = H_t^{1/2}\eta_t$, relationship (2.23) becomes

$$\epsilon_{k,t} = Q'(\mathbf{x})\epsilon_{L,t} \text{ a.s.}, \tag{2.24}$$

where $Q'(\mathbf{x})$ corresponds to a vector containing the coefficients of H_t used for computing the conditional expectation under the gaussian distribution. As H_t is $\mathcal{F}_{t-1}$ measurable, then Q is a function of $\mathbf{x}$. (2.24) means that $\epsilon_{k,t}$ can be written as a linear combination of $\epsilon_{n,t}$, for $n \in L$, given $\mathbf{x}$. If there exists a linear relationship between the components of ϵ_t given $\mathbf{x}$, then the matrix $H_t(\mathbf{x})$ is not a full rank matrix. As $D_t(\mathbf{x})$ is a diagonal matrix, it is always nonsingular, $H_t(\mathbf{x})$ singular implies that $R_t(\mathbf{x})$ is not positive definite. This contradicts $\lambda_1(R_t(\mathbf{x})) > 0$ a.s.. We deduce that

$$\exists \mu > 0, \text{ such that } \forall k, \forall L, k \notin L, \sigma^2_{k|L,t}(\mathbf{x}) \geq \mu \text{ for almost all } \mathbf{x}. \tag{2.25}$$

Consequently, using Assumption 3, we have obtained

$$\begin{aligned} \mathbb{E}\left[g(X_t)|X_{t-1} = \mathbf{x}\right] &\leq 1 + |\mathbf{x}^{\otimes p}|'(T^\star)'M + O\left(\|\mathbf{x}\|^a\right) \\ &\leq g(\mathbf{x}) - |\mathbf{x}^{\otimes p}|'\left(I_N - (T^\star)'\right)M + O\left(\|\mathbf{x}\|^a\right), \end{aligned} \tag{2.26}$$

for all $a > 0$. We denote $N(\mathbf{x}) := \sum_{i=1}^{s} |x_i|^p$. Since $\left(Id - (T^\star)'\right)M > 0$ by Lemma (3.5), then there exists $m_0 > 0$ such that

$$\left(I_s - (T^\star)'\right)M \geq m_0 N(\mathbf{x}), \ \forall \mathbf{x} \in \mathbb{R}^s. \tag{2.27}$$

Similarly, $\exists m_1 > 0$ such that $\forall \mathbf{x} \in \mathbb{R}^s, g(\mathbf{x}) \geq m_1 N(\mathbf{x})$. Using the Hölder's inequality, we have $\forall k \leq p$

$$\sum_{j_1, j_2, \cdots, j_k}^{s} |x_{j_1} x_{j_2} \cdots x_{j_k}| = \left(\sum_{j=1}^{s} |x_j|\right)^k \leq \left(\sum_{j=1}^{s} |x_j|^p\right)^{k/p} s^k. \tag{2.28}$$

Hence using inequality (2.28), $\forall k \leq p, \exists m_2 > 0$ such that

$$g(\mathbf{x}) \leq 1 + \|M\| \sum_{j_1, j_2, \cdots, j_k}^{s} |x_{j_1} x_{j_2} \cdots x_{j_p}| \leq 1 + c_2 N(\mathbf{x}), \tag{2.29}$$

We deduce that

$$\begin{aligned} \mathbb{E}\left[g(X_t)|X_{t-1} = \mathbf{x}\right] &\leq g(\mathbf{x})\left(1 - m_0 \frac{N(\mathbf{x})}{g(\mathbf{x})} + O\left(\frac{N(\mathbf{x})^{a/p}}{g(\mathbf{x})}\right)\right) \\ &\leq g(\mathbf{x})\left(1 - m_0 \frac{N(\mathbf{x})}{1 + m_2 N(\mathbf{x})} + O\left(\frac{N(\mathbf{x})^{a/p}}{m_1 N(\mathbf{x})}\right)\right) \end{aligned} \tag{2.30}$$

We denote $B := \{\mathbf{x} \in \mathbb{R}^s | N(\mathbf{x}) \leq \Gamma\}$, with $\Gamma > 1$. For Γ large enough, $\forall \mathbf{x} \notin B$, and $0 < a < 1$, we have

$$\mathbb{E}\left[g(X_t)|X_{t-1} = \mathbf{x}\right] \leq g(\mathbf{x}) \left(1 - \frac{m_0}{2m_2} + O\left(1\right)\right) < g(\mathbf{x}) \left(1 - \frac{m_0}{3m_2}\right). \tag{2.31}$$

As $1 \leq g(\mathbf{x})$, then $\mathbb{E}\left[g(X_t)|X_{t-1} = \mathbf{x}\right] \leq g(\mathbf{x}) - \varepsilon$, for $\varepsilon > 0$. This proves (2.10), idest $\exists \mu$ a σ-finite invariant measure for $(X_t)_t$ such that $0 < \mu(A) < \infty$.

Now for any $\mathbf{x} \in B$, (2.31) provides

$$\mathbb{E}\left[g(X_t)|X_{t-1} = \mathbf{x}\right] \leq g(\mathbf{x}) + O\left(\|\mathbf{x}\|^a\right) \leq K, \tag{2.32}$$

for some constant $K > 0$. This implies

$$\int_B \left(\int_{B^c} P(\mathbf{x}, y) g(y) d\lambda(y) \right) d\mu(\mathbf{x}) \leq \int_B \mathbb{E}\left[g(X_t)|X_{t-1} = \mathbf{x}\right] d\mu(\mathbf{x}) \leq K\mu(B) \leq \infty. \tag{2.33}$$

Consequently, (2.11) is proved and μ is finite and $\pi = \mu/\mu(E)$ is an invariant probability measure. Then there exists a strictly stationary solution of the stochastic recurrence equation (2.2).

Finally, using inequality (2.31), we obtain (2.12) for $f(\mathbf{x}) = \beta g(\mathbf{x})$, where $\beta \in (0, 1)$. As $m_1 N(\mathbf{x}) \leq g(\mathbf{x})$, then

$$\mathbb{E}_\pi\left[N(X_t)\right] < \infty. \tag{2.34}$$

□

2.3 Uniqueness of stationary Vine-Garch Solutions

Tweedie's criterion provides the existence of an invariant probability measure for Markov chains. However, the uniqueness of such a measure is not ensured. Uniqueness is a significant result as it provides the ergodicity of the stationary solution. This is a significant feature for inference purpose since asymptotic properties for M-estimators are based on Uniform Law of Large Numbers, or the ergodic theorem (see Billingsley, 1995).

Assumption 5. *The sequence of innovations (η_t^*) is strongly stationary.*

Assumption 6. *There exist some strictly positive constant C_h s.t., for any station-*

ary solution, for all t,

$$h_{i|L,t}^{-1} \leq C_h \;\; \mathbb{P}-a.s., \tag{2.35}$$

where $(i|L)$ is associated to an arbitrary node $(i,j|L)$, $L\neq\emptyset$ of the underlying vine $V(n)$.

Note that, when L is empty, the model provides a lower bound for all conditional variances: for every i and t, $h_{i,t}^{-1}\leq C_v$. Let us introduce some intermediate quantities. We denote $C_F>0$ (resp. $C_{\Psi^{-1}}>0$) the Lipschitz constant of $F_{vine}(.)$ (resp. $\Psi^{-1}(.)$). Let us consider two (arbitrarily chosen) stationary solutions (D_t,R_t,ϵ_t) and $(\tilde{D}_t,\tilde{R}_t,\tilde{\varepsilon}_t)$. They share the innovations (η_t^*) and the model parameters. The proof of uniqueness relies on some top Lyapunov exponent of a stochastic matrix process denoted by

$$M_t = \begin{pmatrix} \|\Xi\|_\infty + \|\Lambda\|_\infty \Upsilon_{2,t} & \|\Lambda\|_\infty \Upsilon_{1,t} \\ \Gamma_{2,t} & \Gamma_{1,t} \end{pmatrix}, \tag{2.36}$$

where

$$\left\{\begin{array}{lcl} \Upsilon_{1,t} & = & C_h\sqrt{N}\left(\|D_t\|_s+\|\tilde{D}_t\|_s\right)\{\alpha+\sqrt{N}\|\tilde{D}_t\|_s^2\|\eta_t^*\|_2^2C_h^2\gamma\} \\ \Upsilon_{2,t} & = & C_h\sqrt{N}\left(\|D_t\|_s+\|\tilde{D}_t\|_s\right)\{\beta+\sqrt{N}\|\tilde{D}_t\|_s^2\|\eta_t^*\|_2^2C_h^2\delta\} \\ \gamma & = & C_v^{1/2}N\{\|D_t\|_s+\|\tilde{D}_t\|_s\}\left[1+\dfrac{NC_v\|D_t\|_s^2}{\lambda_1(R_t)}+\dfrac{NC_v\|\tilde{D}_t\|_s^2}{\lambda_1(\tilde{R}_t)}+\dfrac{N^2C_v^2\|D_t\|_s^2\|\tilde{D}_t\|_s^2}{\lambda_1(R_t)\lambda_1(\tilde{R}_t)}\right] \\ \delta & = & \sqrt{N}C_FC_{\Psi^{-1}}\|D_t\|_s\|\tilde{D}_t\|_s\left[1+\dfrac{NC_v\|D_t\|_s^2}{\lambda_1(R_t)}+\dfrac{NC_v\|\tilde{D}_t\|_s^2}{\lambda_1(\tilde{R}_t)}+\dfrac{N^2C_v^2\|D_t\|_s^2\|\tilde{D}_t\|_s^2}{\lambda_1(R_t)\lambda_1(\tilde{R}_t)}\right] \\ \alpha & = & \sqrt{N}C_v^{1/2}\|\eta_t^*\|_s\left\{1+\dfrac{N\|D_t\|_sC_h}{\lambda_1(R_t)}\{\|D_t\|_s+\|\tilde{D}_t\|_s\}\left[1+\dfrac{N\|\tilde{D}_t\|_s^2C_h}{\lambda_1(\tilde{R}_t)}\right]\right\}, \\ \beta & = & \sqrt{N}C_FC_{\Psi^{-1}}\|\tilde{D}_t\|_s\|\eta_t^*\|_s\left\{\dfrac{\sqrt{N}\|D_t\|_s^2C_h}{\lambda_1(R_t)}\left[1+\dfrac{N\|\tilde{D}_t\|_s^2C_h}{\lambda_1(\tilde{R}_t)}\right]+\dfrac{1}{\lambda_1^{1/2}(R_t)+\lambda_1^{1/2}\left(\tilde{R}_t\right)}\right\}. \end{array}\right. \tag{2.37}$$

and

$$\left\{\begin{array}{lcl} \Gamma_{1,t} & = & \|A\|_\infty+N\|B\|_\infty\|\eta_{t-1}^*\|_2^2, \\ \Gamma_{2,t} & = & \|B\|_\infty\|\tilde{D}_{t-1}\|_s^2\dfrac{2\|\eta_t^*\|_2^2}{\lambda_1^{1/2}(R_t)+\lambda_1^{1/2}\left(\tilde{R}_t\right)}NC_FC_{\Psi^{-1}}. \end{array}\right.$$

Assumption 7. *(M_t) is a stationary stochastic process and $\mathbb{E}\left[\log\left(M_t\right)\right]<\infty$ such that its top Lyapunov exponent defined as*

$$\gamma_M := \lim_{t\to\infty}\frac{1}{t}\log\left(M_tM_{t-1}\cdots M_1\right) \tag{2.38}$$

is strictly negative.

Theorem 2.5. *Under Assumptions 1 and 5-7, a strictly stationary solution of the Vine-Garch model is unique and ergodic, given a sequence* $(\eta_t^*)_{t\in\mathbb{Z}}$.

Proof. We remind that $\epsilon_t = D_t u_t = H_t^{1/2}\eta_t$ and $u_t = R_t^{1/2}\eta_t^*$. The model equations define a solution (ϵ_t, D_t, R_t) given (η_t^*). The dynamic system is specified as

$$\begin{cases} \text{Vecd}\left(D_t^2\right) & = & V + A\text{Vecd}\left(D_{t-1}^2\right) + B\vec{\epsilon}_{t-1}, \\ R_t & = & \text{vechof}\left(F_{vine}\left(Pc_t\right)\right), \\ \Psi\left(Pc_t\right) & = & \Omega + \Xi\Psi\left(Pc_{t-1}\right) + \Lambda\zeta_{t-1}. \end{cases} \tag{2.39}$$

A key quantity is the vector of innovations (ζ_t) defined as

$$\begin{cases} \zeta_t & = & \left[v_{i|L,t}v_{j|L,t}\right]_{(i,j|L)\in V(N)}, \\ v_{i|L,t} & = & \dfrac{\epsilon_{i,t} - \mathbb{E}_{t-1}\left[\epsilon_{i,t}|\epsilon_{L,t}\right]}{\sqrt{h_{i|L,t}}}, \end{cases}$$

such that

$$\begin{aligned} h_{i|L,t} &= \text{Var}_{t-1}\left(\epsilon_{i,t}\right) - \text{Cov}_{t-1}\left(\epsilon_{i,t},\epsilon_{L,t}\right)\text{Var}_{t-1}\left(\epsilon_{L,t}\right)^{-1}\text{Cov}_{t-1}\left(\epsilon_{L,t},\epsilon_{i,t}\right), \\ &= e_i'H_te_i - \left(e_i'H_te_L\right).\left(e_L'H_te_L\right)^{-1}.\left(e_L'H_te_i\right). \end{aligned}$$

Above, we have introduced some deterministic matrices (of zeros and ones) e_L s.t. $\epsilon_{L,t} = e_L'\epsilon_t$. The dimension of e_L is $N\times|L|$. More generally, for any $m\times N$-matrix A, Ae_L concatenates the A-columns whose index belongs to L. Using the fact that B is a diagonal matrix and $\epsilon_{i,t} = \sqrt{h_{i,t}}u_{i,t}$, we obtain $\vec{\epsilon}_{i,t} = h_{i,t}u_{i,t}^2$ and

$$\text{Vecd}\left(D_t^2\right) = V + A.\text{Vecd}\left(D_{t-1}^2\right) + B.D_{t-1}^2\vec{u}_{t-1}.$$

where $D_t^2.e = \text{Vecd}\left(D_t^2\right)$.

We first focus on the uniqueness of the conditional variance process. To do so, we consider the difference

$$\begin{aligned} \text{Vecd}\left(D_t^2\right) - \text{Vecd}\left(\tilde{D}_t^2\right) &= A.\left[\text{Vecd}\left(D_{t-1}^2\right) - \text{Vecd}\left(\tilde{D}_{t-1}^2\right)\right] + B.\left[D_t^2\vec{u}_{t-1} - \tilde{D}_{t-1}^2\vec{\tilde{u}}_{t-1}\right], \\ &= A.\left[\text{Vecd}\left(D_{t-1}^2\right) - \text{Vecd}\left(\tilde{D}_{t-1}^2\right)\right] + B.\left[D_{t-1}^2 - \tilde{D}_{t-1}^2\right]\vec{u}_{t-1} \\ &+ B.\tilde{D}_{t-1}^2.\left[\vec{u}_{t-1} - \vec{\tilde{u}}_{t-1}\right]. \end{aligned}$$

Using $D_t^2 \vec{u}_t = \vec{u}_t \odot \operatorname{Vecd}\left(D_{t-1}^2\right)$, we obtain

$$\begin{aligned}
\operatorname{Vecd}\left(D_t^2\right) - \operatorname{Vecd}\left(\tilde{D}_t^2\right) &= A.\left[\operatorname{Vecd}\left(D_{t-1}^2\right) - \operatorname{Vecd}\left(\tilde{D}_{t-1}^2\right)\right] \\
&+ \; B\vec{u}_{t-1} \odot \left[\operatorname{Vecd}\left(D_{t-1}^2\right) - \operatorname{Vecd}\left(\tilde{D}_{t-1}^2\right)\right] + B\left[\vec{u}_{t-1} - \vec{\tilde{u}}_{t-1}\right] \odot \operatorname{Vecd}\left(\tilde{D}_{t-1}^2\right).
\end{aligned}$$

Furthermore

$$\begin{aligned}
\vec{u}_t - \vec{\tilde{u}}_t &= (u_t - \tilde{u}_t) \odot (u_t + \tilde{u}_t) \\
&= (u_t + \tilde{u}_t) \odot \left(R_t^{1/2} - \tilde{R}_t^{1/2}\right) \eta_t^*.
\end{aligned}$$

Using the spectral norm, the previous quantity can be upper bounded as

$$\|\vec{u}_t - \vec{\tilde{u}}_t\|_s \leq \|u_t + \tilde{u}_t\|_\infty . \| \left(R_t^{1/2} - \tilde{R}_t^{1/2}\right) \eta_t^* \|_s.$$

Since $\|\eta_t^*\|_s = \|\eta_t^*\|_2$ (as for any vector), note that

$$\|u_t\|_\infty = \|R_t^{1/2}\eta_t^*\|_\infty \leq \|R_t^{1/2}\eta_t^*\|_s \leq \|R_t\|_s^{1/2}\|\eta_t^*\|_2 \leq \sqrt{N}\|\eta_t^*\|_2.$$

Using theorem 6.2 of Hingham (2008), for any unitarily invariant norm $\|\cdot\|$, we have

$$\|R_t^{1/2} - \tilde{R}_t^{1/2}\| \leq \frac{1}{\lambda_1^{1/2}\left(R_t\right) + \lambda_1^{1/2}\left(\tilde{R}_t\right)}\|R_t - \tilde{R}_t\|.$$

Recall that the norm $\|.\|$ is unitarily invariant if $\|UAV\| = \|A\|$ for all matrix A and all unitary matrices U and V, ie $UU' = Id$ and $VV' = Id$. For instance, the spectral norm $\|A\|_s = \rho\left(A'A\right)^{1/2} = \lambda_{max}\left(A\right)$ satisfies

$$\|UAV\|_s = \rho\left((UAV)' .UAV\right)^{1/2} = \rho\left(V'A'AV\right)^{1/2} = \rho\left(A'A\right)^{1/2} = \|A\|_s,$$

and is then unitarily invariant. Hence

$$\begin{aligned}
\| \left(R_t^{1/2} - \tilde{R}_t^{1/2}\right) \eta_t^* \|_s &\leq \|R_t^{1/2} - \tilde{R}_t^{1/2}\|_s \|\eta_t^*\|_s \\
&\leq \frac{1}{\lambda_1^{1/2}\left(R_t\right) + \lambda_1^{1/2}\left(\tilde{R}_t\right)}\|R_t - \tilde{R}_t\|_s \|\eta_t^*\|_2.
\end{aligned}$$

Besides,

$$\begin{aligned}
\|R_t - \tilde{R}_t\|_s &\leq \sqrt{N}\|R_t - \tilde{R}_t\|_\infty \\
&\leq \sqrt{N} C_F \|Pc_t - \tilde{P}c_t\|_\infty \\
&\leq \sqrt{N} C_F C_{\Psi^{-1}} \|\Psi\left(Pc_t\right) - \Psi\left(\tilde{P}c_t\right)\|_\infty.
\end{aligned}$$

As B and A are diagonal matrices, their spectral norms are equal to their infinite norm. We obtain the upper bound

$$\begin{aligned}
\|\text{Vecd}\left(D_t^2\right) - \text{Vecd}\left(\tilde{D}_t^2\right)\|_s &\leq \|A\|_s\|\text{Vecd}\left(D_{t-1}^2\right) - \text{Vecd}\left(\tilde{D}_{t-1}^2\right)\|_s \\
&+ \|B\|_s\|\vec{u}_{t-1}\|_s\|\text{Vecd}\left(D_{t-1}^2\right) - \text{Vecd}\left(\tilde{D}_{t-1}^2\right)\|_s \\
&+ \|B\|_s\|\text{Vecd}\left(\tilde{D}_{t-1}^2\right)\|_\infty\|u_t + \tilde{u}_t\|_\infty\|\eta_t^*\|_2\frac{\|R_t - \tilde{R}_t\|_s}{\lambda_1^{1/2}\left(R_t\right) + \lambda_1^{1/2}\left(\tilde{R}_t\right)} \\
&\leq \Gamma_{1,t}\|\text{Vecd}\left(D_{t-1}^2\right) - \text{Vecd}\left(\tilde{D}_{t-1}^2\right)\|_s + \Gamma_{2,t}\|\Psi\left(Pc_t\right) - \Psi\left(\tilde{P}c_t\right)\|_\infty,
\end{aligned} \tag{2.40}$$

where

$$\begin{cases}
\Gamma_{1,t} = \|A\|_\infty + N\|B\|_\infty\|\eta_{t-1}^*\|_2^2, \\
\Gamma_{2,t} = \|B\|_\infty\|\tilde{D}_{t-1}\|_s^2\dfrac{2\|\eta_t^*\|_2^2}{\lambda_1^{1/2}\left(R_t\right) + \lambda_1^{1/2}\left(\tilde{R}_t\right)}NC_FC_{\Psi^{-1}}.
\end{cases}$$

We now focus on the uniqueness of the partial correlation process. We consider the difference

$$\Psi\left(Pc_t\right) - \Psi\left(\tilde{P}c_t\right) = \Xi\left(\Psi\left(Pc_{t-1}\right) - \Psi\left(\tilde{P}c_{t-1}\right)\right) + \Lambda\left(\zeta_{t-1} - \tilde{\zeta}_{t-1}\right).$$

In this framework, Ξ and Λ are parameterized as diagonal matrices. We have

$$\|\Psi\left(Pc_t\right) - \Psi\left(\tilde{P}c_t\right)\|_\infty \leq \|\Xi\|_\infty\|\Psi\left(Pc_{t-1}\right) - \Psi\left(\tilde{P}c_{t-1}\right)\|_\infty + \|\Lambda\|_\infty\|\zeta_{t-1} - \tilde{\zeta}_{t-1}\|_\infty. \tag{2.41}$$

The quantity of interest is the vector of innovations, that is

$$v_{ij|L,t} - \tilde{v}_{ij|L,t} = \frac{r_{i|L,t}r_{j|L,t}}{\sqrt{h_{i|L,t}}\sqrt{h_{j|L,t}}} - \frac{\tilde{r}_{i|L,t}\tilde{r}_{j|L,t}}{\sqrt{\tilde{h}_{i|L,t}}\sqrt{\tilde{h}_{j|L,t}}}, \tag{2.42}$$

where, using the Gaussian assumption, we have

$$\begin{aligned}
r_{i|L,t} &= \epsilon_{i,t} - \mathbb{E}_{t-1}\left[\epsilon_{i,t}|\epsilon_{L,t}\right] \\
&= \epsilon_{i,t} - \left(e_i'H_te_L\right).\left(e_L'H_te_L\right)^{-1}\epsilon_{L,t} \\
&= \left[e_i' - \left(e_i'H_te_L\right).\left(e_L'H_te_L\right)^{-1}e_L'\right]\epsilon_t \\
&:= e_i'\mathbf{p}_L\left(\epsilon_t\right).
\end{aligned} \tag{2.43}$$

Here, $\mathbf{p}_L\left(\cdot\right)$ is the projector on the orthogonal of the subspace $< H_te_L >$ in $\mathbb{R}^N$, relatively to the H_t^{-1}-euclidian norm, defined by $\|\boldsymbol{x}\|_H = \boldsymbol{x}'H_t^{-1}\boldsymbol{x}$ [2]. By decomposing

[2] Indeed, if $\boldsymbol{x}_j = H_te_Lg_j$ for any $|L|\times 1$-vector $g_j = [\delta_{i,j}]_{j=1,\dots,|L|}$, we check that $\mathbf{p}_L(\boldsymbol{x}_j) = 0$. Moreover, when a vector $\boldsymbol{v}$ belongs to $< H_te_L >^\perp$, then $\boldsymbol{v}'H_t^{-1}H_te_Lg_j = \boldsymbol{v}'e_Lg_j = 0$ for every j, i.e. $\boldsymbol{v}'e_L = 0$. This implies $\mathbf{p}_L(\boldsymbol{v}) = \boldsymbol{v}$.

the projector $\mathbf{p}_L$ in its canonical space, we see that $\|\mathbf{p}_L\|_s = 1$ obviously. Similarly, $\|\tilde{\mathbf{p}}_L\|_s = 1$.

Recall that $\epsilon_t = D_t R_t^{1/2}\eta_t^*$. Using the same steps as in (2.43), we obtain

$$\tilde{r}_{i|L,t} = e_i'\tilde{\mathbf{p}}_L\left(\tilde{\epsilon}_t\right), \ \tilde{\epsilon}_t = \tilde{D}_t\tilde{R}_t^{1/2}\eta_t^*.$$

Now we have

$$\|\zeta_{t-1} - \tilde{\zeta}_{t-1}\|_\infty = \sup_{(i,j|L)} |v_{ij|L,t} - \tilde{v}_{ij|L,t}|,$$

which implies we need to control $|r_{i|L,t} - \tilde{r}_{i|L,t}|$ and $|h_{i|L,t} - \tilde{h}_{i|L,t}|$.

Step 1. We have

$$\begin{aligned} r_{i|L,t} - \tilde{r}_{i|L,t} &= e_i'\mathbf{p}_L\left(\epsilon_t\right) - e_i'\tilde{\mathbf{p}}_L\left(\tilde{\epsilon}_t\right) \\ &= e_i'\left[\mathbf{p}_L - \tilde{\mathbf{p}}_L\right]\left(\epsilon_t\right) + e_i'\tilde{\mathbf{p}}_L\left(\epsilon_t - \tilde{\epsilon}_t\right) \end{aligned}$$

We obtain

$$\begin{aligned} |r_{i|L,t} - \tilde{r}_{i|L,t}| &\leq \|(\mathbf{p}_L - \tilde{\mathbf{p}}_L)(\epsilon_t)\|_\infty + \|\tilde{\mathbf{p}}_L\left(\epsilon_t - \tilde{\epsilon}_t\right)\|_\infty \\ &\leq \|(\mathbf{p}_L - \tilde{\mathbf{p}}_L)(\epsilon_t)\|_2 + \|\tilde{\mathbf{p}}_L\left(\epsilon_t - \tilde{\epsilon}_t\right)\|_2. \end{aligned}$$

Note that, for any vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|_H^2 = \boldsymbol{x}H_t^{-1}\boldsymbol{x} \geq \boldsymbol{x}'\boldsymbol{x}/\rho(H_t)$. Since $\rho(H_t) \leq Tr(H_t) \leq \sum_{j=1}^N h_{j,t} \leq N\|D_t\|_s^2$. Therefore, we get

$$\|\boldsymbol{x}\|_2 \leq \sqrt{N}\|D_t\|_s\|\boldsymbol{x}\|_H.$$

Moreover, for every vector $\boldsymbol{x}$, $\|\boldsymbol{x}\|_H^2 = \boldsymbol{x}'H_t^{-1}\boldsymbol{x} \leq \|\boldsymbol{x}\|_2^2\|H_t^{-1}\|_s$ (diagonalize H_t in an orthonormal basis). This means

$$\|\boldsymbol{x}\|_H \leq C_v^{1/2}\lambda_1(R_t)^{-1/2}\|\boldsymbol{x}\|_2. \tag{2.44}$$

Since the spectral norm is the matrix norm that is associated to the usual euclidian norm $\|\cdot\|_2$, we have

$$\begin{aligned} |r_{i|L,t} - \tilde{r}_{i|L,t}| &\leq \|(\mathbf{p}_L - \tilde{\mathbf{p}}_L)(\epsilon_t)\|_2 + \|\tilde{\mathbf{p}}_L\left(\epsilon_t - \tilde{\epsilon}_t\right)\|_2 \\ &\leq \|(\mathbf{p}_L - \tilde{\mathbf{p}}_L)\|_s\|\epsilon_t\|_s + \|\tilde{\mathbf{p}}_L\|_s\|\epsilon_t - \tilde{\epsilon}_t\|_2 \\ &\leq \|(\mathbf{p}_L - \tilde{\mathbf{p}}_L)\|_s\|\epsilon_t\|_2 + \|\epsilon_t - \tilde{\epsilon}_t\|_2. \end{aligned}$$

Furthermore,

$$\begin{aligned}
&\tilde{\mathbf{p}}_L - \mathbf{p}_L = -\left(I_N - H_t e_L \left(e_L' H_t e_L\right)^{-1} e_L'\right) + \left(I_N - \tilde{H}_t e_L \left(e_L' \tilde{H}_t e_L\right)^{-1} e_L'\right) \\
&= \left(H_t - \tilde{H}_t\right) e_L \left(e_L' H_t e_L\right)^{-1} e_L' + \tilde{H}_t e_L \left[\left(e_L' H_t e_L\right)^{-1} - \left(e_L' \tilde{H}_t e_L\right)^{-1}\right] e_L' \\
&= \left[\left(D_t - \tilde{D}_t\right) R_t D_t + \tilde{D}_t \left(R_t - \tilde{R}_t\right) D_t + \tilde{D}_t \tilde{R}_t \left(D_t - \tilde{D}_t\right)\right] e_L \left(e_L' H_t e_L\right)^{-1} e_L' \\
&+ \tilde{H}_t e_L \left(e_L' H_t e_L\right)^{-1} \left[\left(e_L' \tilde{H}_t e_L\right) - \left(e_L' H_t e_L\right)\right] \left(e_L' \tilde{H}_t e_L\right)^{-1} e_L'.
\end{aligned}$$

Note that $\|(e_L' H_t e_L)^{-1}\|_s$ is the inverse of the smallest eigenvalue of $e_L' H_t e_L$. By the Courant-Raleigh theorem, $\lambda_1(e_L' H_t e_L)$ is larger than $\lambda_1(H_t)$. Then, $\|(e_L' H_t e_L)^{-1}\|_s \leq \lambda_1(H_t)^{-1} = \|H_t^{-1}\|_s$. Since $H_t^{-1} = D_t^{-1} R_t^{-1} D_t^{-1}$, we obtain

$$\|(e_L' H_t e_L)^{-1}\|_s \leq \|H_t^{-1}\|_s \leq \|D_t^{-1}\|_s^2 \|R_t^{-1}\|_s \leq C_v \lambda_1(R_t)^{-1}.$$

Moreover, it is easy to check that $\|e_L\|_s = \|e_L'\|_s = 1$. Since

$$\|D_t - \tilde{D}_t\|_s \leq \max_i |h_{i,t} - \tilde{h}_{i,t}| / (h_{i,t}^{1/2} + \tilde{h}_{i,t}^{1/2}) \leq C_v^{1/2} \|\text{Vecd}\left(D_t^2\right) - \text{Vecd}\left(\tilde{D}_t^2\right)\|_s,$$

we have

$$\begin{aligned}
\|\mathbf{p}_L - \tilde{\mathbf{p}}_L\|_s &\leq \{\|D_t - \tilde{D}_t\|_s \|R_t\|_s \|D_t\|_s + \|\tilde{D}_t\|_s \|R_t - \tilde{R}_t\|_s \|D_t\|_s + \|\tilde{D}_t\|_s \|\tilde{R}_t\|_s \|D_t - \tilde{D}_t\|_s\} \\
&\cdot \quad \|(e_L' H_t e_L)^{-1}\|_s \left(\|1 + \|\tilde{H}_t\|_s \|(e_L' \tilde{H}_t e_L)^{-1}\|_s\right) \\
&\leq \{C_v^{1/2} \|\text{Vecd}\left(D_t^2\right) - \text{Vecd}\left(\tilde{D}_t^2\right)\|_s N (\|D_t\|_s + \|\tilde{D}_t\|_s) \\
&+ \sqrt{N} C_F C_{\Psi^{-1}} \|\Psi\left(P c_t\right) - \Psi\left(\tilde{P} c_t\right)\|_\infty \|D_t\|_s \|\tilde{D}_t\|_s\} \\
&\cdot \quad C_h \lambda_1(R_t)^{-1} \left(1 + N \|\tilde{D}_t\|_s^2 C_h \lambda_1(\tilde{R}_t)^{-1}\right).
\end{aligned}$$

We also have

$$\begin{aligned}
&\|\epsilon_t\|_2 \leq \|D_t\|_s \|R_t^{1/2}\|_s \|\eta_t^*\|_2 \\
&\quad \leq \quad \|D_t\|_s \lambda_{max}^{1/2}\left(R_t\right) \|\eta_t^*\|_2 \leq \|D_t\|_s \sqrt{N} \|\eta_t^*\|_2.
\end{aligned}$$

Moreover,

$$\begin{aligned}
\|\epsilon_t - \tilde{\epsilon}_t\|_s &\leq \|\left(D_t R_t^{1/2} - \tilde{D}_t \tilde{R}_t^{1/2}\right) \eta_t^*\|_s \\
&\leq \|D_t - \tilde{D}_t\|_s \|R_t^{1/2}\|_s \|\eta_t^*\|_s + \|\tilde{D}_t\|_s \|R_t^{1/2} - \tilde{R}_t^{1/2}\|_s \|\eta_t^*\|_s \\
&\leq \left[C_v^{1/2} \|\text{Vecd}\left(D_t^2\right) - \text{Vecd}\left(\tilde{D}_t^2\right)\|_s \sqrt{N} + \|\tilde{D}_t\|_s K \|\Psi\left(P c_t\right) - \Psi\left(\tilde{P} c_t\right)\|_\infty\right] \|\eta_t^*\|_s, \\
\text{where } K &= \frac{1}{\lambda_1^{1/2}\left(R_t\right) + \lambda_1^{1/2}\left(\tilde{R}_t\right)} \sqrt{N} C_F C_{\Psi^{-1}}.
\end{aligned}$$

Consequently, for every (i, L) deduced from the vine structure, we obtain

$$|r_{i|L,t} - \tilde{r}_{i|L,t}| \leq \alpha \|\text{Vecd}\left(D_t^2\right) - \text{Vecd}\left(\tilde{D}_t^2\right)\|_s + \beta \|\Psi\left(Pc_t\right) - \Psi\left(\tilde{P}c_t\right)\|_\infty. \quad (2.45)$$

where

$$\begin{cases} \alpha &= \sqrt{N} C_v^{1/2} \|\eta_t^*\|_s \left\{ 1 + \dfrac{N\|D_t\|_s C_h}{\lambda_1(R_t)} \{\|D_t\|_s + \|\tilde{D}_t\|_s\} \left[1 + \dfrac{N\|\tilde{D}_t\|_s^2 C_h}{\lambda_1(\tilde{R}_t)} \right] \right\}, \\ \beta &= \sqrt{N} C_F C_{\Psi^{-1}} \|\tilde{D}_t\|_s \|\eta_t^*\|_s \left\{ \dfrac{\sqrt{N}\|D_t\|_s^2 C_h}{\lambda_1(R_t)} \left[1 + \dfrac{N\|\tilde{D}_t\|_s^2 C_h}{\lambda_1(\tilde{R}_t)} \right] + \dfrac{1}{\lambda_1^{1/2}\left(R_t\right) + \lambda_1^{1/2}\left(\tilde{R}_t\right)} \right\}. \end{cases}$$

Step 2. We now focus on the discrepancy $|h_{i|L,t} - \tilde{h}_{i|L,t}|$. We have

$$\begin{aligned} h_{i|L,t} - \tilde{h}_{i|L,t} &= e_i'\left(H_t - \tilde{H}_t\right) e_i - e_i'\left(H_t - \tilde{H}_t\right) e_L \left(e_L' H_t e_L\right)^{-1} \left(e_L' H_t e_i\right) \\ &+ e_i' \tilde{H}_t e_L \left(e_L' H_t e_L\right)^{-1} \left[e_L' \tilde{H}_t e_L - e_L' H_t e_L\right] \left(e_L' \tilde{H}_t e_L\right)^{-1} \left(e_L' H_t e_i\right) \\ &+ e_i' \tilde{H}_t e_L \left(e_L' \tilde{H}_t e_L\right)^{-1} e_L' \left(H_t - \tilde{H}_t\right) e_i, \end{aligned}$$

which implies

$$\begin{aligned} |h_{i|L,t} - \tilde{h}_{i|L,t}| &\leq \|H_t - \tilde{H}_t\|_s \Big[1 + C_v \lambda_1(R_t)^{-1} \|H_t\|_s + C_v \lambda_1(\tilde{R}_t)^{-1} \|\tilde{H}_t\|_s \\ &+ C_v^2 \lambda_1(R_t)^{-1} \lambda_1(\tilde{R}_t)^{-1} \|H_t\|_s \|\tilde{H}_t\|_s \Big] \\ &\leq \Big(C_v^{1/2} \|\text{Vecd}\left(D_t^2\right) - \text{Vecd}\left(\tilde{D}_t^2\right)\|_s N \{\|D_t\|_s + \|\tilde{D}_t\|_s\} \\ &+ \sqrt{N} C_F C_{\Psi^{-1}} \|D_t\|_s \|\tilde{D}_t\|_s \|\Psi\left(Pc_t\right) - \Psi\left(\tilde{P}c_t\right)\|_\infty \Big) \\ &\cdot \left[1 + \frac{N C_v \|D_t\|_s^2}{\lambda_1(R_t)} + \frac{N C_v \|\tilde{D}_t\|_s^2}{\lambda_1(\tilde{R}_t)} + \frac{N^2 C_v^2 \|D_t\|_s^2 \|\tilde{D}_t\|_s^2}{\lambda_1(R_t)\lambda_1(\tilde{R}_t)} \right] \\ &\leq \gamma \|\text{Vecd}\left(D_t^2\right) - \text{Vecd}\left(\tilde{D}_t^2\right)\|_s + \delta \|\Psi\left(Pc_t\right) - \Psi\left(\tilde{P}c_t\right)\|_\infty, \qquad (2.46) \end{aligned}$$

where

$$\begin{cases} \gamma &= C_v^{1/2} N \{\|D_t\|_s + \|\tilde{D}_t\|_s\} \left[1 + \dfrac{N C_v \|D_t\|_s^2}{\lambda_1(R_t)} + \dfrac{N C_v \|\tilde{D}_t\|_s^2}{\lambda_1(\tilde{R}_t)} + \dfrac{N^2 C_v^2 \|D_t\|_s^2 \|\tilde{D}_t\|_s^2}{\lambda_1(R_t)\lambda_1(\tilde{R}_t)} \right] \\ \delta &= \sqrt{N} C_F C_{\Psi^{-1}} \|D_t\|_s \|\tilde{D}_t\|_s \left[1 + \dfrac{N C_v \|D_t\|_s^2}{\lambda_1(R_t)} + \dfrac{N C_v \|\tilde{D}_t\|_s^2}{\lambda_1(\tilde{R}_t)} + \dfrac{N^2 C_v^2 \|D_t\|_s^2 \|\tilde{D}_t\|_s^2}{\lambda_1(R_t)\lambda_1(\tilde{R}_t)} \right] \end{cases}$$

Consequently, we obtain the following relationship for (2.42)

$$\begin{aligned} v_{i,j|L,t} - \tilde{v}_{i,j|L,t} &= \frac{\left(r_{i|L,t} - \tilde{r}_{i|L,t}\right) r_{j|L,t}}{\sqrt{h_{i|L,t}}\sqrt{h_{j|L,t}}} + \frac{\tilde{r}_{i|L,t}\left(r_{j|L,t} - \tilde{r}_{j|L,t}\right)}{\sqrt{h_{i|L,t}}\sqrt{h_{j|L,t}}} \\ &+ \tilde{r}_{i|L,t}\tilde{r}_{j|L,t}\{\frac{1}{\sqrt{h_{i|L,t}}\sqrt{h_{j|L,t}}} - \frac{1}{\sqrt{\tilde{h}_{i|L,t}}\sqrt{\tilde{h}_{j|L,t}}}\}. \end{aligned}$$

For any (i, L) we consider, $h_{i|L,t} \leq \|D_t\|_s^2$ everywhere, because the variance of a residual is smaller than the variance of any random variable. Therefore, we get

$$\begin{aligned} |\frac{1}{\sqrt{h_{i|L,t}}\sqrt{h_{j|L,t}}} - \frac{1}{\sqrt{\tilde{h}_{i|L,t}}\sqrt{\tilde{h}_{j|L,t}}}| &\leq \frac{C_h^2}{\sqrt{\tilde{h}_{i|L,t}}\sqrt{\tilde{h}_{j|L,t}} + \sqrt{\tilde{h}_{i|L,t}}\sqrt{\tilde{h}_{j|L,t}}}\{h_{i|L,t}h_{j|L,t} - \tilde{h}_{i|L,t}\tilde{h}_{j|L,t}\} \\ &\leq C_h^3\left[\left(h_{i|L,t} - \tilde{h}_{i|L,t}\right)h_{j|L,t} + \tilde{h}_{i|L,t}\left(h_{j|L,t} - \tilde{h}_{j|L,t}\right)\right] \\ &\leq C_h^3\{\|D_t\|_s^2|h_{i|L,t} - \tilde{h}_{i|L,t}| + \|\tilde{D}_t\|_s^2|h_{j|L,t} - \tilde{h}_{j|L,t}|\}, \end{aligned} \quad (2.47)$$

and

$$|r_{i|L,t}| \leq \|\mathbf{p}_L(\epsilon_t)\|_\infty \leq \|\mathbf{p}_L(\epsilon_t)\|_2 \leq \|\mathbf{p}_L\|_s.\|\epsilon_t\|_2 \leq \|\epsilon_t\|_2 \leq \sqrt{N}\|D_t\|_s\|_s\|\eta_t^*\|_2. \quad (2.48)$$

Consequently, using (2.46), (2.47) and (2.48), (2.42) can be upper bounded as

$$\begin{aligned} &|v_{i,j|L,t} - \tilde{v}_{i,j|L,t}| \leq C_h\sqrt{N}\left(\|D_t\|_s + \|\tilde{D}_t\|_s\right) \\ &\cdot \quad \left\{\left(\alpha\|\text{Vecd}\left(D_t^2\right) - \text{Vecd}\left(\tilde{D}_t^2\right)\|_s + \beta\|\Psi\left(Pc_t\right) - \Psi\left(\tilde{P}c_t\right)\|_\infty\right)\right. \\ &+ \quad \left.\sqrt{N}\|\tilde{D}_t\|_s^2\|\eta_t^*\|_2^2C_h^2\left(\gamma\|\text{Vecd}\left(D_t^2\right) - \text{Vecd}\left(\tilde{D}_t^2\right)\|_s + \delta\|\Psi\left(Pc_t\right) - \Psi\left(\tilde{P}c_t\right)\|_\infty\right)\right\}. \end{aligned}$$

Hence using the previous inequality, we obtain

$$\|\zeta_t - \tilde{\zeta}_t\|_\infty \leq \Upsilon_{1,t}\|\text{Vecd}\left(D_t^2\right) - \text{Vecd}\left(\tilde{D}_t^2\right)\|_s + \Upsilon_{2,t}\|\Psi\left(Pc_t\right) - \Psi\left(\tilde{P}c_t\right)\|_\infty, \quad (2.49)$$

with

$$\left\{\begin{aligned} \Upsilon_{1,t} &= C_h\sqrt{N}\left(\|D_t\|_s + \|\tilde{D}_t\|_s\right)\{\alpha + \sqrt{N}\|\tilde{D}_t\|_s^2\|\eta_t^*\|_2^2C_h^2\gamma\} \\ \Upsilon_{2,t} &= C_h\sqrt{N}\left(\|D_t\|_s + \|\tilde{D}_t\|_s\right)\{\beta + \sqrt{N}\|\tilde{D}_t\|_s^2\|\eta_t^*\|_2^2C_h^2\delta\} \end{aligned}\right.$$

Using (2.49) and (2.41), we have

$$\begin{aligned}\|\Psi\left(Pc_t\right)-\Psi\left(\tilde{P}c_t\right)\|_\infty &\leq \{\|\Xi\|_\infty+\|\Lambda\|_\infty\Upsilon_{2,t}\}\|\Psi\left(Pc_{t-1}\right)-\Psi\left(\tilde{P}c_{t-1}\right)\|_\infty \\ &+ \|\Lambda\|_\infty\Upsilon_{1,t}\|\mathrm{Vecd}\left(D_t^2\right)-\mathrm{Vecd}\left(\tilde{D}_t^2\right)\|_s.\end{aligned} \tag{2.50}$$

We denote

$$\|\mu_t\|=\begin{pmatrix}\|\Psi\left(Pc_{t-1}\right)-\Psi\left(\tilde{P}c_{t-1}\right)\|_\infty \\ \|\mathrm{Vecd}\left(D_t^2\right)-\mathrm{Vecd}\left(\tilde{D}_t^2\right)\|_s\end{pmatrix},\ M_t=\begin{pmatrix}\|\Xi\|_\infty+\|\Lambda\|_\infty\Upsilon_{2,t} & \|\Lambda\|_\infty\Upsilon_{1,t} \\ \Gamma_{2,t} & \Gamma_{1,t}\end{pmatrix}.$$

Using (2.40) and (2.50), we deduce that

$$\begin{aligned}\|\mu_t\| &\leq M_t\|\mu_{t-1}\| \\ &\leq \{\prod_{k=0}^{t-p}M_{t-k}\}\|\mu_{t-p-1}\|,\end{aligned}$$

for any $p\in\mathbb{N}$. Under Assumption 7, $\lim_{p\to\infty}\|M_tM_{t-1}\cdots M_{t-p}\|=0$ $\mathbb{P}-$a.s., for a fixed t using Lemma 2.1 of Francq and Zakoian (2010). We deduce that $\mu_t \underset{t\to\infty}{\longrightarrow} 0$. This implies that $\Psi(Pc_t)=\Psi(\tilde{P}c_t)$ a.s. and $D_t=\tilde{D}_t$ a.s., which then implies $R_t=\tilde{R}_t$ a.s. and $\epsilon_t=\tilde{\epsilon}_t$ a.s.. This concludes the proof of uniqueness. Furthermore, ergodicity is obtained as a consequence of corollary 7.17 in Douc and al. (2014).

A sufficient condition for uniqueness is that the top Lyapunov exponent γ_M is strictly negative. This condition holds if $\mathbb{E}\left[\log\left(\|M_t\|\right)\right]<0$.

□

3 Asymptotic Properties

We provided conditions for the existence and the uniqueness of strictly stationary solutions of the Vine-GARCH model. These results are significant for asymptotic properties since law of large numbers and central limit theorems can be applied.

We consider the DGP given in (1.1). From an inference point of view, a practical issue arises when computing log-likelihoods, which is the choice of some initial values to generate the sequences (D_t), (R_t) and then (H_t), $t = 1, \ldots, T$. Given some fixed values for ϵ_0, D_0 and R_0, we obtain log-likelihoods. In this Section only, the latter log-likelihoods will be denoted by $\widetilde{QL}_{1,T}(\theta_v; \epsilon)$ and $\widetilde{QL}_{2,T}(\theta_v, \theta_c; \epsilon)$. More generally, all quantities with a " ˜ " are deduced from the process with fixed arbitrary starting values at $t = 0$. Therefore, they are distinct from the "theoretical" log-likelihoods $QL_{1,T}(\theta_v; \epsilon)$ and $QL_{2,T}(\theta_v, \theta_c; \epsilon)$, for which the initial values are coming from the stationary laws [3]. Actually, this subtlety has no consequence because we will assume irrelevance of initial values: see Assumptions 10 and 18 and Section A.

3.1 Consistency

To show the weak consistency, we need a set of assumptions given as follows.

Assumption 8. *The variance parameters θ_v (resp. correlation parameters θ_c) belong to a compact set Θ_v in $\mathbb{R}_+^{3N}$ (resp. Θ_c in in $\mathbb{R}_+^{2N^2(N-1)^2/4+N(N-1)/2}$). The true parameter $\theta_0 = (\theta_{0,v}, \theta_{0,c})'$ belongs to the interior of the compact set $\Theta := \Theta_v \times \Theta_c$.*

Assumption 9. *The sequence of innovations (η_t) is strongly stationary, $\mathbb{E}_{t-1}[\eta_t] = 0$, $\mathbb{E}_{t-1}[\eta_t \eta_t'] = I_N$ and $\mathbb{E}_{t-1}[\eta_{i,t}|\eta_{j,t}] = 0$ when $i \neq j$. Moreover, $\eta_t \in \mathbb{R}^N$ has a nondegenerate distribution.*

Assumption 10. *The initial values are asymptotically irrelevant, which means*

$$\sup_{\theta \in \Theta} |QL_{2,T}(\theta; \epsilon) - \widetilde{QL}_{2,T}(\theta; \epsilon)| = \circ_p(1).$$

Moreover and as expected, we need the classic assumptions that guarantee the strong consistency of univariate GARCH(1,1) QML estimates.

[3] Equivalently, they can be seen as coming from a stationary solution $(\epsilon_t)_{t \in \mathbb{Z}}$.

Assumption 11. . *Let $M_0 = (M_{0,t})_t$ be a sequence of matrix defined by $M_{0,t} = \left(\eta_t^2, 1\right)' (\kappa, \tau)$, where the index i are removed for clarity. Then $\gamma_{M_0} < 0$, the top Lyapunov exponent, defined as*

$$\gamma_{M_0} := \inf_{t \in \mathbb{N}^\star} \frac{1}{t} \mathbb{E}\left[\log\left(\|M_{0,t} M_{0,t-1} \cdots M_{0,1}\|\right)\right] = \lim_{t \to \infty} \frac{1}{t} \log\left(\|M_{0,t} M_{0,t-1} \cdots M_{0,1}\|\right) \ a.s.$$

Besides, $\forall \theta_v \in \Theta_v$, $\tau < 1$.

Assumption 12. *Let (A_t, B_t) defined as*

$$\begin{aligned} A_t &:= \sup_{\theta : \|\theta_v - \theta_{0,v}\| < \alpha} \| \left(\nabla \Psi(Pc_t)\right)^{-1} \Lambda \nabla_{D_t} \zeta(Pc_t, D_t, \epsilon_t) \nabla_{\theta_v} D_t \|, \\ B_t &:= \sup_{\theta : \|\theta_v - \theta_{0,v}\| < \alpha} \| \left(\nabla \Psi(Pc_t)\right)^{-1} \left[\Xi \nabla \Psi(Pc_t) + \Lambda \nabla_{Pc_t} \zeta(Pc_t, D_t, \epsilon_t)\right] \|. \end{aligned}$$

For some $\alpha > 0$, the stochastic matrix process (A_t, B_t) is stationary, $E[A_t] < +\infty$ and

$$\sum_{k \geq 1} E\left[B_{t-1} B_{t-2} \cdots B_{t-k} A_{t-k-1}\right] < \infty.$$

Theorem 3.1. *Let $\hat{\theta}_T = \left(\hat{\theta}_{T,v}, \hat{\theta}_{T,c}\right)'$ a sequence of pseudo-maximum likelihood estimators verifying (2.10) and (2.11). Then under Assumptions 4 and 8-12,*

$$\hat{\theta}_T \xrightarrow{\mathbb{P}} \theta_0 \quad when \quad T \to \infty.$$

We shall proceed step-by-step to prove the weak consistency of the two-step estimator. We denote $\theta_{0\backslash c} = (\theta_{0,v}, \theta_c)$. The next three steps shall be demonstrated successively.

1. Identifiability of the parameters, which can be expressed in our framework as

$$\{\forall t \in \mathbb{Z},\ D_t(\theta_v) = D_t(\theta_{0,v}) \text{ and } R_t(\theta) = R_t(\theta_0)\ \mathbb{P}_{\theta_0} \text{ as}\} \Rightarrow \theta = \theta_0.$$

2. The optimum θ_0 is well-separated: if $\hat{\theta}_{T,v} \to \theta_{0,v}$ a.s., and $\|\theta_c - \theta_{0,c}\| > \gamma$, for some $\gamma > 0$ then $l_{2,t}(\epsilon_t; \theta_{0,v}, \theta_{0,c}) \in L^1(\mathbb{R})$ and

$$\mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t; \theta_{0,v}, \theta_c)\right] > \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t; \theta_0)\right].$$

3. Let $\Theta_{0\backslash c} = \{\theta = (\theta_{0,v}, \theta_c) \in \Theta\} = \{\theta_{0,v}\} \times \Theta_c$. For every $\theta^* \in \Theta_{0\backslash c}$ with $\|\theta_c^* - \theta_{0,c}\| > 0$ and every $\pi > 0$, there exists an open ball $V(\theta^*, \pi)$ around θ^*

in the space $\Theta_{0\backslash c}$ s.t.

$$\mathbb{E}_{\theta_0}\left[\inf_{\theta\in V(\theta^*,\pi)} l_{2,t}(\epsilon_t;\theta)\right] \geq \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta^*)\right] - \pi.$$

Note that a proof of the strong consistency of $\hat{\theta}_{T,v}$ can be found in Francq and Zakoian (2004). In their proof, similar steps as above (and Assumption 10) are demonstrated for $QL_{1,T}(\epsilon;\theta_v)$, the first step likelihood.

Remark about Assumption 10. The asymptotic irrelevance of the initial values is probably the most difficult step to prove for weak consistency. To do so, when comparing $QL_{2,T}(\theta;\epsilon)$ and $\widetilde{QL}_{2,T}(\theta;\epsilon)$, we need to evaluate the rate of convergence of $\|R_t - \tilde{R}_t\|$. For the sake of clarity, we assume that this assumption holds. A detailed proof is considered secondary w.r.t. the core of the consistency result, but it can be found in Section A.

We need the following lemma, whose proof is postponed after the proof of Theorem 3.1.

Lemma 3.2. *Under Assumptions 8-11,*

$$\sup_{\theta\in\Theta} |QL_{2,T}(\hat{\theta}_{T,v},\theta_c) - QL_{2,T}(\theta_{0,v},\theta_c)| = o_P(1).$$

Proof. Step 1. We now prove the identifiability part of the vine-Garch model. Due to the identifiability of GARCH models, when $D_t(\theta_v) = D_t(\theta_{0,v})$ for every t and almost everywhere, this means that $\theta_v = \theta_{0,v}$.

Let us state the identifiability of the correlation-related parameters. To do so, we define $\mathcal{D}_\theta(z) = \Lambda z$ and $\mathcal{Q}_\theta(z) = I_N - \Xi z$. There is a one-to-one relationship between the components of the lower (or upper) triangular part of $R_t(\theta)$, and $P_{C_t}(\theta)$, the vector of partial correlations, through $F_{vine}(.)$. Then $R_t(\theta) = R_t(\theta_0)$ $\mathbb{P}_{\theta_0}$ a.s. implies $P_{C_t}(\theta) = P_{C_t}(\theta_0)$ $\mathbb{P}_{\theta_0}$ a.s.. For a given sequence of innovations (η_t), we write the partial correlation dynamics $\Psi\left(Pc_t(\theta)\right) = \Omega + \Xi\Psi\left(Pc_{t-1}(\theta)\right) + \Lambda\zeta_{t-1}(\theta)$ as

$$\begin{aligned}
\mathcal{Q}_\theta(B)\Psi\left(Pc_t(\theta)\right) &= \Omega + \mathcal{D}_\theta(B)\zeta_t(\theta) \\
\Leftrightarrow \Psi\left(Pc_t(\theta)\right) &= \mathcal{Q}_\theta^{-1}(B)\left[\Omega + \mathcal{D}_\theta(B)\zeta_t(\theta)\right] \\
\Leftrightarrow \Psi\left(Pc_t(\theta)\right) &= \mathcal{Q}_\theta^{-1}(B)\mathcal{D}_\theta(B)\left[\mathcal{D}_\theta^{-1}(1)\Omega + \zeta_t(\theta)\right].
\end{aligned}$$

Now, for some θ and θ_0, suppose

$$\mathcal{Q}_\theta^{-1}(B)\mathcal{D}_\theta(B)\left[\mathcal{D}_\theta^{-1}(1)\Omega + \zeta_t(\theta)\right] = \mathcal{Q}_{\theta_0}^{-1}(B)\mathcal{D}_{\theta_0}(B)\left[\mathcal{D}_{\theta_0}^{-1}(1)\Omega_0 + \zeta_t(\theta_0)\right] \text{ a.s.} \forall t.$$

This means

$$\mathcal{Q}_\theta^{-1}(B)\mathcal{D}_\theta(B)\zeta_t(\theta) - \mathcal{Q}_{\theta_0}^{-1}(B)\mathcal{D}_{\theta_0}(B)\zeta_t(\theta_0) = \mathcal{Q}_{\theta_0}^{-1}(1)\Omega_0 - \mathcal{Q}_\theta^{-1}(1)\Omega \text{ a.s.},$$

which can be rewritten as

$$(I_N - \Xi B)^{-1}\Lambda B\zeta_t(\theta) - (I_N - \Xi_0 B)^{-1}\Lambda_0 B\zeta_t(\theta_0) = \mathcal{Q}_{\theta_0}^{-1}(1)\Omega_0 - \mathcal{Q}_{\theta_0}^{-1}(1)\Omega.$$

The latter is equivalent to

$$\sum_{k\geq 0}\left[\Xi^k B^{k+1}\Lambda\zeta_t(\theta) - \Xi_0^k B^{k+1}\Lambda_0\zeta_t(\theta_0)\right] = \mathcal{Q}_{\theta_0}^{-1}(1)\Omega_0 - \mathcal{Q}_\theta^{-1}(1)\Omega,$$

$$\Leftrightarrow \sum_{k\geq 0}\left[\Xi^k\Lambda\zeta_{t-k-1}(\theta) - \Xi_0^k\Lambda_0\zeta_{t-k-1}(\theta_0)\right] = \mathcal{Q}_{\theta_0}^{-1}(1)\Omega_0 - \mathcal{Q}_\theta^{-1}(1)\Omega,$$

implying

$$\Lambda\zeta_{t-1}(\theta) - \Lambda_0\zeta_{t-1}(\theta_0) + \Xi\Lambda\zeta_{t-2}(\theta) - \Xi_0\Lambda_0\zeta_{t-2}(\theta_0) + \cdots = \mathcal{Q}_{\theta_0}^{-1}(1)\Omega_0 - \mathcal{Q}_\theta^{-1}(1)\Omega, \quad (3.1)$$

for every t and almost everywhere. Hence, a.s., it is equivalent to

$$\Lambda\zeta_{t-1}(\theta) - \Lambda_0\zeta_{t-1}(\theta_0) = M_{t-2}, \quad (3.2)$$

where M_{t-2} is a random variable that is measurable wrt $\sigma(\eta_{t-s}, s \geq 2)$. For an arbitrary parameter θ, let $\epsilon_t = H_t^{1/2}(\theta)\eta_t$, that depends implicitly on the underlying parameter. Let us consider an element of $\zeta_{t-1}(\theta)$, which corresponds to a conditioned set, say i, j, and a conditioning set, say L. Then, we have

$$\zeta_{t-1}^{(ij|L)}(\theta) = \frac{\epsilon_{i,t-1} - \mathbb{E}\left[\epsilon_{i,t-1}|\epsilon_{L,t-1}, \mathcal{F}_{t-2}\right]}{\sqrt{h_{i|L,t-1}}} \cdot \frac{\epsilon_{j,t-1} - \mathbb{E}\left[\epsilon_{j,t-1}|\epsilon_{L,t-1}, \mathcal{F}_{t-2}\right]}{\sqrt{h_{j|L,t-1}}}.$$

Both denominators are $\mathcal{F}_{t-2}$ measurable and depend on θ through the variance and correlation processes. We rewrite $\epsilon_{i,t} = e_i'(H_t^{1/2}(\theta)\eta_t)$, with $e_i = (0, \cdots, 0, 1, 0, \cdots, 0)'$ with 1 at the ith component. Focusing on one of the numerators, the quantity of interest is

$$
\begin{aligned}
\epsilon_{i,t-1} - \mathbb{E}\left[\epsilon_{i,t-1}|\epsilon_{L,t-1},\mathcal{F}_{t-2}\right] &= e_i'H_{t-1}^{1/2}(\theta)\eta_{t-1} - e_i'H_{t-1}^{1/2}(\theta)\mathbb{E}\left[\eta_{t-1}|\epsilon_{L,t-1},\mathcal{F}_{t-2}\right] \\
&= e_i'H_{t-1}^{1/2}(\theta)\{\eta_{t-1} - \mathbb{E}\left[\eta_{t-1}|\epsilon_{L,t-1},H_{t-1}(\theta)\right]\} \\
&= e_i'H_{t-1}^{1/2}(\theta)\{\eta_{t-1} - \mathbb{E}\left[\eta_{t-1}|f_L(\eta_{t-1},\theta)\right]\}.
\end{aligned}
$$

f_L is a linear function of the components of η_{t-1} and depends on the specific conditioning set of the vine. It also depends on θ because the linear coefficients of f_L are the sub-components of $H_t^{1/2}(\theta)$. Now, for all t,

$$
\zeta_{t-1}^{(ij|L)}(\theta) = \text{Trace}\left(e_i'H_{t-1}^{1/2}(\theta)\alpha_{t-1}^L(\theta)\alpha_{t-1}^{L\prime}(\theta)H_{t-1}^{1/2}(\theta)e_j\right) / \left(\sqrt{h_{i|L,t-1}(\theta)}\sqrt{h_{j|L,t-1}(\theta)}\right),
$$

with $\alpha_{t-1}^L(\theta) = \eta_{t-1} - \mathbb{E}\left[\eta_{t-1}|f_L(\eta_{t-1},\theta)\right]$.

But we assumed $R_t(\theta) = R_t(\theta_0)$, $D_t(\theta) = D_t(\theta_0)$ for t. Consequently, $H_t(\theta) = H_t(\theta_0)$, the observations ϵ_t are the same under $\mathbb{P}_\theta$ and $\mathbb{P}_{\theta_0}$, and $\alpha_t^L(\theta) = \alpha_t^L(\theta_0)$. This implies that $\zeta_t(\theta) = \zeta_t(\theta_0)$ and (3.2) becomes componentwise

$$
\begin{aligned}
&\lambda_{(ij|L)}\zeta_{t-1}^{(ij|L)}(\theta) - \lambda_{0,(ij|L)}\zeta_{t-1}^{(ij|L)}(\theta_0) = m_{t-2} \\
&\Leftrightarrow (\lambda_{(ij|L)} - \lambda_{0,(ij|L)})e_i'H_{t-1}^{1/2}(\theta)\alpha_{t-1}^L(\theta)(\alpha_{t-1}^L(\theta))'H_{t-1}^{1/2}(\theta)e_j = m_{t-2}
\end{aligned}
$$

for some $\mathcal{F}_{t-2}-$measurable function m_{t-2}. The l.h.s. corresponds to a quadratic form of (η_{t-1}), whose coefficients are some functions of $H_{t-1}(\theta)$. This can be rewritten

$$
\begin{aligned}
& (\lambda_{(ij|L)} - \lambda_{0,(ij|L)})\sum_{k,l=1}^N H_{i,k,t-1}^{1/2}(\theta)H_{j,k,t-1}^{1/2}(\theta)\alpha_{k,t-1}^L(\theta)\alpha_{l,t-1}^L(\theta) = m_{t-2} \\
\Leftrightarrow\ & (\lambda_{(ij|L)} - \lambda_{0,(ij|L)})(\mu_{11}\eta_{1,t-1}^2 + \mu_{12}\eta_{1,t-1}\eta_{2,t-1} + \cdots + a_1\eta_{1,t-1} + \cdots + \mu_{kk}\eta_{k,t-1}^2 + \mu_{k1}\eta_{k,t-1}\eta_{1,t-1} \\
& + \cdots + a_k\eta_{k,t-1} + \cdots + a_N\eta_{N,t-1} + C) = m_{t-2},
\end{aligned} \tag{3.3}
$$

for some $\mathcal{F}_{t-2}$-measurable coefficients $\mu_{i,j}$, a_k and a constant C. Taking the conditional expectation $\mathbb{E}\left[.|\eta_{-k,t-1},\mathcal{F}_{t-2}\right]$, with $\eta_{-k,t-1}$ the vector η_{t-1} excluding $\eta_{k,t-1}$, using the assumption $\mathbb{E}[\eta_{t-1}\,|\,\eta_{-k,t-1},\mathcal{F}_{t-2}] = 0$ (c.f. Assumption 9), and substracting to (3.3), we obtain

$$
(\lambda_{(ij|L)} - \lambda_{0,(ij|L)})(\mu_{kk}\eta_{k,t-1}^2 + a_k\eta_{k,t-1} - \mu_{kk}\mathbb{E}_{t-2}\left[\eta_{k,t-1}^2\right]) = 0, \text{ a.s.} \tag{3.4}
$$

If $\lambda_{(ij|L)} \neq \lambda_{0,(ij|L)}$ in (3.4), then a solution is $\eta_{k,t-1} \in \{\alpha,\beta\}$. This contradicts Assumption 9. Consequently, $\lambda_{(ij|L)} = \lambda_{0,(ij|L)}$. This holds for all the components of $\zeta_{t-1}(\theta)$, hence $\Lambda = \Lambda_0$. Plugging this last inequality in (3.1), we obtain

$$(\Xi - \Xi_0)\Lambda\zeta_{t-2}(\theta) = M_{t-3}. \tag{3.5}$$

The same steps can be applied as previously

$$\lambda_{(ij|L)}(\xi_{(ij|L)} - \xi_{0,(ij|L)})(\mu_{11}\eta_{1,t-2}^2 + \mu_{12}\eta_{1,t-2}\eta_{2,t-2} + \cdots + a_1\eta_{1,t-2} + \cdots + \mu_{kk}\eta_{k,t-2}^2 + \mu_{k1}\eta_{k,t-2}\eta_{1,t-2}$$
$$+ \cdots + a_k\eta_{k,t-2} + + \cdots a_N\eta_{N,t-2} + C) = m_{t-3}.$$

Taking the conditional expectation $\mathbb{E}\left[.|\eta_{-k,t-2}\right]$ and using the same steps as previously, we obtain $\xi_{(ij|L)} = \xi_{0,(ij|L)}$, hence $\Xi = \Xi_0$. Finally, a.s.

$$\sum_{k\geq 0}\left[\Xi^k B^{k+1}\Lambda\zeta_t(\theta) - \Xi_0^k B^{k+1}\Lambda_0\zeta_t(\theta_0)\right] = \mathcal{Q}_{\theta_0}^{-1}(1)\Omega_0 - \mathcal{Q}_{\theta_0}^{-1}(1)\Omega$$
$$\Leftrightarrow \quad 0 = \mathcal{Q}_{\theta_0}^{-1}(1)\Omega_0 - \mathcal{Q}_{\theta}^{-1}(1)\Omega$$

As $\mathcal{Q}_{\theta_0}(1) = \mathcal{Q}_{\theta}(1)$, this implies $\Omega = \Omega_0$. □

Proof. Step 2. We now show that the limit criterion is minimized at the true value. It is important to note that the second step is conditional to the first step estimator, idest we deal with $l_{2,t}(\epsilon_t; \hat{\theta}_{T,v}, \theta_c)$.

For all $\theta \in \Theta$,

$$\mathbb{E}_{\theta_0}\left[l_{2,t}^-(\epsilon_t;\theta)\right] \leq \mathbb{E}_{\theta_0}\left[\log^-(|R_t|)\right] \leq \max\left(0, -\log(|R_t|)\right) < \infty, \tag{3.6}$$

by Assumption 4. Consequently, $\mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta)\right]$ belongs to $\mathbb{R} \cup \{+\infty\}$. Now at the true parameter value, we show $\mathbb{E}_{\theta_0}\left[|l_{2,t}(\epsilon_t;\theta_0)|\right] < \infty$.

Indeed, the determinant of $R_t(\theta_0)$ is bounded from above by $\text{Trace}(R_t)^N$, i.e. N^N. Therefore, without any assumption,

$$\mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_0)\right] = \mathbb{E}_{\theta_0}\left[\log(|R_t(\theta_0)|)\right] + \text{Trace}\left(\mathbb{E}_{\theta_0}\left[\eta_t'\eta_t\right]\right) \leq N\log N + N.$$

Therefore, we obtain that $l_{2,t}(\epsilon_t;\theta_0) \in L^1$.

Denoting by $\alpha_{i,t}$ the eigenvalues of $R_t(\theta_0)R_t^{-1}(\theta_{0\backslash c})$, which are positive, $\theta_{0\backslash c} =$

$(\theta_{0,v}, \theta_c)$ and $u_t = D_t(\theta_{0,v})^{-1}\epsilon_t$, we have

$$\begin{aligned}
&\mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_{0\backslash c}) - l_{2,t}(\epsilon_t;\theta_0)\right] \\
&= \mathbb{E}_{\theta_0}\left[\log(|R_t(\theta_{0\backslash c})||R_t^{-1}(\theta_0)|)\right] + \mathbb{E}_{\theta_0}\left[u_t'\left(R_t^{-1}(\theta_{0\backslash c}) - R_t^{-1}(\theta_0)\right)u_t\right] \\
&= \mathbb{E}_{\theta_0}\left[\log(|R_t(\theta_{0\backslash c})||R_t^{-1}(\theta_0)|)\right] + \mathbb{E}_{\theta_0}\left[\eta_t' R_t^{1/2}(\theta_0)'\left(R_t^{-1}(\theta_{0\backslash c}) - R_t^{-1}(\theta_0)\right)R_t^{1/2}(\theta_0)\eta_t\right] \\
&= \mathbb{E}_{\theta_0}\left[\log(|R_t(\theta_{0\backslash c})||R_t^{-1}(\theta_0)|)\right] + \mathbb{E}_{\theta_0}\left[\text{Trace}\left(\eta_t'\left(R_t^{1/2}(\theta_0)'R_t^{-1}(\theta_{0\backslash c})R_t^{1/2}(\theta_0) - I_N\right)\eta_t\right)\right] \\
&= \mathbb{E}_{\theta_0}\left[\log(|R_t(\theta_{0\backslash c})||R_t^{-1}(\theta_0)|)\right] + \mathbb{E}_{\theta_0}\left[\text{Trace}\left(\left(R_t^{1/2}(\theta_0)'R_t^{-1}(\theta_{0\backslash c})R_t^{1/2}(\theta_0) - I_N\right)\eta_t\eta_t'\right)\right] \\
&= \mathbb{E}_{\theta_0}\left[\log(|R_t(\theta_{0\backslash c})||R_t^{-1}(\theta_0)|)\right] + \mathbb{E}_{\theta_0}\left[\text{Trace}\left(\left(R_t^{1/2}(\theta_0)'R_t^{-1}(\theta_{0\backslash c})R_t^{1/2}(\theta_0) - I_N\right)\mathbb{E}_{t-1}[\eta_t\eta_t']\right)\right] \\
&= \mathbb{E}_{\theta_0}\left[\log(|R_t(\theta_{0\backslash c})||R_t^{-1}(\theta_0)|)\right] + \mathbb{E}_{\theta_0}\left[\text{Trace}\left(R_t(\theta_0)R_t^{-1}(\theta_{0\backslash c}) - I_N\right)\right] \\
&= \mathbb{E}_{\theta_0}\left[\sum_{i=1}^N(\alpha_{i,t} - 1 - \log(\alpha_{i,t}))\right] \geq 0,
\end{aligned} \tag{3.7}$$

because $\forall x > 0$, $\log(x) \leq x - 1$. The inequality $\log(x) \leq x - 1$ holds if and only if $x = 1$. In our case, that means $\alpha_{i,t} = 1$, $\forall i$, which is $R_t(\theta_{0\backslash c}) = R_t(\theta_0)$ $\mathbb{P}_{\theta_0}$ a.s.. By stationarity, this reasoning can be made at time $t-1$, which would give $R_{t-1}(\theta_{0\backslash c}) = R_{t-1}(\theta_0)$ $\mathbb{P}_{\theta_0}$ a.s.. Hence for any t, the relationship $R_t(\theta_{0\backslash c}) = R_t(\theta_0)$ $\mathbb{P}_{\theta_0}$ a.s. holds by stationarity. By step 1, this means $\theta_0 = \theta_{0\backslash c}$. □

Proof. Step 3. For a given $\theta^* \in \Theta_{0\backslash v}$, $\theta_c^* \neq \theta_{0,c}$, consider a sequence of open balls of radius $1/k$, $k \in \mathbb{N}$ defined by $V_k(\theta^*) := \{\theta \in \Theta_{0\backslash v} | \, \|\theta - \theta^*\| \leq 1/k\}$. Since the sequence of random variable $(\inf_{\theta \in V_k(\theta^*)} l_{2,t}(\epsilon_t;\theta))_k$ is increasing, the Beppo-Levi Theorem applies:

$$\lim_{k\to\infty} \mathbb{E}_{\theta_0}\left[\inf_{\theta \in V_k(\theta^*)} l_{2,t}(\epsilon_t;\theta)\right] = \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta^*)\right],$$

providing the result. □

Proof. Theorem (3.1). Under our assumptions, $\hat{\theta}_{T,v}$ converges weakly to $\theta_{0,v}$ (see Theorem 7.1 in Francq and Zakoian, 2010, e.g.). Now, let us prove the weak convergence of $\hat{\theta}_{T,c}$ to $\theta_{0,c}$, that is

$$\forall \alpha > 0, \; \lim_{T\to\infty} \mathbb{P}\left(\|\hat{\theta}_{T,c} - \theta_{0,c}\| > \alpha\right) = 0. \tag{3.8}$$

By Assumption 8, Θ and then $\Theta_{0\backslash c}$ are compact sets. For any given $\pi > 0$ and for every $\theta^* \in \Theta_{0\backslash c}$, $\theta^* \neq \theta_0$ with $\|\theta_c^* - \theta_{0,c}\| \geq \alpha/2$, let us associate an open ball $U(\theta^*) \subset \Theta_{0\backslash c}$ s.t.

$$\mathbb{E}_{\theta_0}\left[\inf_{\theta \in U(\theta^*)} l_{2,t}(\epsilon_t;\theta)\right] \geq \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta^*)\right] - \pi.$$

We know it is always possible due to the previous Step 3. Since the function $\theta \in \Theta_{0\backslash c} \to \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_{0,v},\theta_c)\right] - \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_0)\right]$ is strictly positive (c.f. Step 2) and continuous on the compact subset $\mathcal{E}_0(\alpha) := \{\theta \in \Theta_{0\backslash c} \,|\, \|\theta_c - \theta_{0,c}\| \geq \alpha/2\}$, it reaches its minimum $2\mu > 0$. Therefore, for any given $\theta^* \in \mathcal{E}_0(\alpha)$, set $\pi := \pi(\theta^*) = \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta^*)\right] - \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_0)\right] - \mu > 0$.

Moreover, set $U(\theta_0) := \{\theta \in \Theta_{0\backslash c} : \|\theta - \theta_0\| < \alpha\}$. Then

$$\Theta_{0\backslash c} \subset U(\theta_0) \cup \bigcup_{\theta \in \mathcal{E}_0(\alpha)} U(\theta).$$

Since $\Theta_{0\backslash c}$ can be covered by a finite set of open balls, there is a finite set of points $\theta_1, \ldots, \theta_n$ in $\mathcal{E}_0(\alpha)$ s.t.

$$\Theta_{0\backslash c} \subset U(\theta_0) \cup \bigcup_{i=1,\ldots,n} U(\theta_i).$$

Equation (3.8) becomes

$$\mathbb{P}\left(\|\hat{\theta}_{T,c} - \theta_{0,c}\| > \alpha\right) \leq \mathbb{P}\left((\theta_{0,v}, \hat{\theta}_{T,c}) \in \bigcup_{i=1,\ldots,n} U(\theta_i)\right) \leq \sum_{i=1,\ldots,n} \mathbb{P}\left((\theta_{0,v}, \hat{\theta}_{T,c}) \in U(\theta_i)\right).$$

By definition of $\hat{\theta}_T$, we obtain for all $i = 1, \cdots, n$

$$\begin{aligned}
&\mathbb{P}\left((\theta_{0,v}, \hat{\theta}_{T,c}) \in U(\theta_i)\right) \leq \mathbb{P}\left(\inf_{\theta \in U(\theta_i)} \widetilde{QL}_{2,T}(\theta;\epsilon) \leq \widetilde{QL}_{2,T}(\theta_{0,v}, \hat{\theta}_{T,c};\epsilon)\right) \\
&\leq \mathbb{P}\left(\inf_{\theta \in U(\theta_i)} QL_{2,T}(\theta;\epsilon) \leq QL_{2,T}(\hat{\theta}_{T,v}, \hat{\theta}_{T,c};\epsilon) + 2\sup_{\theta \in \Theta}|QL_{2,T}(\theta;\epsilon) - \widetilde{QL}_{2,T}(\theta;\epsilon)|\right. \\
&+ \left.|QL_{2,T}(\theta_{0,v}, \hat{\theta}_{T,c};\epsilon) - QL_{2,T}(\hat{\theta}_T;\epsilon)|\right) \\
&\leq \mathbb{P}\left(\inf_{\theta \in U(\theta_i)} QL_{2,T}(\theta;\epsilon) \leq QL_{2,T}(\hat{\theta}_{T,v}, \theta_{0,c};\epsilon)| + 2\sup_{\theta \in \Theta}|QL_{2,T}(\theta;\epsilon) - \widetilde{QL}_{2,T}(\theta;\epsilon)|\right. \\
&+ \left.|QL_{2,T}(\theta_{0,v}, \hat{\theta}_{T,c};\epsilon) - QL_{2,T}(\hat{\theta}_T;\epsilon)|\right) \\
&\leq \mathbb{P}\left(\inf_{\theta \in U(\theta_i)} QL_{2,T}(\theta;\epsilon) \leq \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_0)\right] + 2\sup_{\theta \in \Theta}|QL_{2,T}(\theta;\epsilon) - \widetilde{QL}_{2,T}(\theta;\epsilon)|\right. \\
&+ \left.|QL_{2,T}(\theta_0;\epsilon) - \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_0)\right]| + |QL_{2,T}(\theta_{0,v}, \hat{\theta}_{T,c};\epsilon) - QL_{2,T}(\hat{\theta}_T;\epsilon)|\right) \\
&\leq \mathbb{P}\left(\mathbb{E}_{\theta_0}\left[\inf_{\theta \in U(\theta_i)} l_{2,t}(\epsilon_t;\theta)\right] \leq \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_0)\right] + 2\sup_{\theta \in \Theta}|QL_{2,T}(\theta;\epsilon) - \widetilde{QL}_{2,T}(\theta;\epsilon)| + |\mathcal{R}_{\theta_i}|\right. \\
&+ \left.|QL_{2,T}(\theta_0;\epsilon) - \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_0)\right]| + |QL_{2,T}(\theta_{0,v}, \hat{\theta}_{T,c};\epsilon) - QL_{2,T}(\hat{\theta}_T;\epsilon)|\right), \qquad (3.9)
\end{aligned}$$

where $\mathcal{R}_{\theta_i} = \frac{1}{T}\sum_{t=1}^{T} \inf_{\theta\in U(\theta_i)} l_{2,t}(\epsilon_t;\theta) - \mathbb{E}_{\theta_0}\left[\inf_{\theta\in U(\theta_i)} l_{2,t}(\epsilon_t;\theta)\right]$. Invoking step 3 and the way the neighborhoods have been built, for any $i = 1,\ldots,n$,

$$\mathbb{E}_{\theta_0}\left[\inf_{\theta\in U(\theta_i)} l_{2,t}(\epsilon_t;\theta)\right] \geq \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_0)\right] + \mu.$$

Using the property $\{X+Y \leq a+b\} \subset \{X \leq a\} \cup \{Y \leq b\}$, $a,b \geq 0$ and X,Y any random variables, (3.9) becomes

$$\begin{aligned}
&\mathbb{P}\left((\theta_{0,v},\hat{\theta}_{c,T}) \in U(\theta_i)\right) \leq \mathbb{P}\Big(\mu \leq 2\sup_{\theta\in\Theta}|QL_{2,T}(\theta;\epsilon) - \widetilde{QL}_{2,T}(\theta;\epsilon)| + |\mathcal{R}_{\theta_i}| \\
&+ \quad |QL_{2,T}(\theta_0;\epsilon) - \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_0)\right]| + |QL_{2,T}(\theta_{0,v},\hat{\theta}_{T,c};\epsilon) - QL_{2,T}(\hat{\theta}_T;\epsilon)|\Big) \\
&\leq \quad \mathbb{P}\left(\frac{\mu}{4} < 2\sup_{\theta\in\Theta}|QL_{2,T}(\theta;\epsilon) - \widetilde{QL}_{2,T}(\theta;\epsilon)|\right) + \mathbb{P}\left(\frac{\mu}{4} < |QL_{2,T}(\theta_0;\epsilon) - \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_0)\right]|\right) \\
&+ \quad \mathbb{P}\left(\frac{\mu}{4} < |\mathcal{R}_{\theta_i}|\right) + \mathbb{P}\left(\frac{\mu}{4} < |QL_{2,T}(\theta_{0,v},\hat{\theta}_{T,c};\epsilon) - QL_{2,T}(\hat{\theta}_T;\epsilon)|\right). \qquad (3.10)
\end{aligned}$$

Under Assumption 10, the initial values generating the process are asymptotically irrelevant. For some $\delta > 0$ and $T > T_1$, this implies

$$\mathbb{P}\left(\frac{\mu}{4} < 2\sup_{\theta\in\Theta}|QL_{2,T}(\theta;\epsilon) - \widetilde{QL}_{2,T}(\theta;\epsilon)|\right) < \delta/4. \qquad (3.11)$$

As for the second probability of the r.h.s. in (3.10), we use the ergodic theorem of Billingsley (1995), and for $T > T_2$, we obtain

$$\mathbb{P}\left(\frac{\mu}{4} < |QL_{2,T}(\theta_0;\epsilon) - \mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta_0)\right]|\right) < \delta/4. \qquad (3.12)$$

Let us focus on the the third term in the r.h.s. Although the quantity $l_{2,t}(\epsilon_t;\theta)$ is not necessarily integrable, the ergodic theorem can still be used as $\mathbb{E}_{\theta_0}\left[l_{2,t}(\epsilon_t;\theta)\right] \in \mathbb{R}\cup\{\infty\}$. Furthermore, $l_{2,t}(\epsilon_t;\theta)$ is a measurable function of an ergodic process, hence, as in Exercise 7.4 in Francq and Zakoian (2010), the ergodic theorem of Billingsley (1995) can be applied to $(\inf_{\theta\in U(\theta_i)} l_{2,t}(\epsilon_t;\theta))_t$ as follows

$$\liminf_{T\to\infty} \frac{1}{T}\sum_{t=1}^{T} \inf_{\theta\in U(\theta_i)} l_{2,t}(\epsilon_t;\theta) = \mathbb{E}_{\theta_0}\left[\inf_{\theta\in U(\theta_i)} l_{2,t}(\epsilon_t;\theta)\right].$$

Plugging this convergence result into (3.10), for $\delta > 0,\ T > T_3$, we obtain

$$\mathbb{P}\left(\frac{\mu}{4} < |\mathcal{R}_{\theta_i}|\right) < \delta/4. \tag{3.13}$$

Note that the derivative of $\theta_v \mapsto QL_{2,T}(\theta_v, \theta_{0,c}; \epsilon)$ is uniformly bounded under Assumption 4 (recall the arguments in the proof of Step 2). Invoking Lemma 3.2, we can tackle the fourth term of (3.10): if $t > T_4$, we have

$$\mathbb{P}\left(\frac{\mu}{4} < |QL_{2,T}(\theta_0; \epsilon) - QL_{2,T}(\hat{\theta}_{T,v}, \theta_{0,c}; \epsilon)|\right) < \delta/4. \tag{3.14}$$

Consequently, with (3.11), (3.12), (3.13) and (3.14), for $T > T_1 \vee T_2 \vee T_3 \vee T_4$, (3.10) becomes

$$\mathbb{P}\left(\hat{\theta}_T \in U(\theta_i)\right) \leq \delta. \tag{3.15}$$

Since δ can be chosen arbitrarily small, this proves the convergence in probability of $\left(\hat{\theta}_{T,v}, \hat{\theta}_{T,c}\right)'$ to the true parameter vector θ_0. □

Proof. Lemma 3.2. Applying a Taylor expansion to $QL_{2,T}(\hat{\theta}_{T,v}, \theta_c; \epsilon)$ around $\theta_{0,v}$, we obtain

$$\frac{1}{T}\sum_{t=1}^{T} l_{2,t}(\epsilon_t; \hat{\theta}_{T,v}, \theta_c) = \frac{1}{T}\sum_{t=1}^{T} l_{2,t}(\epsilon_t; \theta_{0,v}, \theta_c) + (\hat{\theta}_{T,v} - \theta_{0,v})\frac{1}{T}\sum_{t=1}^{T} \nabla_{\theta_v} l_{2,t}(\epsilon_t; \bar{\theta}_v, \theta_c),$$

for some $\bar{\theta}_v$, $\|\bar{\theta}_v - \theta_{0,v}\| < \|\theta_{0,v} - \hat{\theta}_{T,v}\|$. Using the consistency of $\hat{\theta}_{T,v}$, it is sufficient to prove that

$$\frac{1}{T}\sum_{t=1}^{T} \sup_{\{\theta \in \Theta \,|\, \|\theta_v - \theta_{0,v}\| < \alpha\}} \|\nabla_{\theta_v} l_{2,t}(\epsilon_t; \theta_v, \theta_c)\| = O_P(1), \tag{3.16}$$

for some (small) $\alpha > 0$. Applying some matrix derivation rules (see Lütkepohl, 1996), the analytical score of the second step likelihood with respect to the i-th element of θ_v is given by

$$\begin{aligned}
\partial_{\theta_v^i} l_{2,t}(\epsilon_t; \theta) &= \partial_{\theta_v^i}\left[\log(|R_t|) + \epsilon_t' D_t^{-1} R_t^{-1} D_t^{-1} \epsilon_t\right] \\
&= \text{Trace}\left(R_t^{-1}(\partial_{\theta_v^i} R_t)\right) + \text{Trace}\left(\epsilon_t \epsilon_t' \partial_{\theta_v^i}\left[D_t^{-1} R_t^{-1} D_t^{-1}\right]\right) \\
&= \text{Trace}\left(R_t^{-1}(\partial_{\theta_v^i} R_t)\right) - \text{Trace}\left(\epsilon_t \epsilon_t'\left[D_t^{-1}(\partial_{\theta_v^i} D_t) D_t^{-1} R_t^{-1} D_t^{-1}\right]\right) \\
&- \text{Trace}\left(\epsilon_t \epsilon_t'\left[D_t^{-1} R_t^{-1}(\partial_{\theta_v^i} R_t) R_t^{-1} D_t^{-1}\right]\right) - \text{Trace}\left(\epsilon_t \epsilon_t'\left[D_t^{-1} R_t^{-1} D_t^{-1}(\partial_{\theta_v^i} D_t) D_t^{-1}\right]\right).
\end{aligned}$$

Obviously, the matrices D_t^{-1} are bounded from above by positive constants due to the definition of our univariate GARCH dynamics. Concerning correlations, we know that R_t^{-1} is bounded from above, due to Assumption 4. As for the derivatives of R_t, note that $\|\nabla_{\theta_v} R_t\| \leq \|\nabla F_{vine}(Pc_t).\nabla_{\theta_v} Pc_t\|$ and that the derivative of $F_{vine}(\cdot)$ is bounded a.e. under the latter assumption.

Consequently, there exists some positive constant C such that, for any $\alpha > 0$,

$$\sup_{\theta:\|\theta_v-\theta_{0,v}\|<\alpha} |\nabla_{\theta_v} l_{2,t}(\epsilon_t;\theta_c,\theta_v)| \leq C. \sup_{\theta:\|\theta_v-\theta_{0,v}\|<\alpha} \{(\|\nabla_{\theta_v} D_t\|+\|\nabla_{\theta_v} Pc_t\|)\|\epsilon_t\|^2+\|\nabla_{\theta_v} Pc_t\|\}.$$

Let us focus on $\nabla_{\theta_v} Pc_t$. By the chain rule, we have

$$\begin{aligned}\nabla_{\theta_v} Pc_t &= (\nabla\Psi(Pc_{t-1}))^{-1}\left[\Xi\nabla\Psi(Pc_{t-1}) + \Lambda\nabla_{Pc}\zeta(Pc_{t-1},D_{t-1},\epsilon_{t-1})\right]\nabla_{\theta_v} Pc_{t-1}\\ &+ (\nabla\Psi(Pc_{t-1}))^{-1}\Lambda\nabla_D\zeta(Pc_{t-1},D_{t-1},\epsilon_{t-1})\nabla_{\theta_v} D_{t-1},\end{aligned}$$

and then

$$\begin{aligned}\sup_{\theta:\|\theta_v-\theta_{0,v}\|<\alpha} \|\nabla_{\theta_v} Pc_t\| &\leq A_{t-1} + B_{t-1} \sup_{\theta:\|\theta_v-\theta_{0,v}\|<\alpha} \|\nabla_{\theta_v} Pc_{t-1}\|\\ &\leq A_{t-1} + \sum_{k=1}^{\infty} B_{t-1}B_{t-2}\cdots B_{t-k}A_{t-k-1}. \qquad (3.17)\end{aligned}$$

Assumption 12 provides sufficient conditions so that the latter series belongs to L^1. As a consequence, the existence of the series (3.17) is ensured a.s. But we need a stronger assumption than in Theorem 1.1. of Bougerol and Picard (1992) typically, because of the integrability requirement. This implies

$$\frac{1}{T}\sum_{t=1}^{T} \sup_{\theta:\|\theta_v-\theta_{0,v}\|<\alpha} \|\nabla_{\theta_v} Pc_t\|.(\|\epsilon_t\|^2+1) = O_P(1).$$

We now focus on $\|\nabla_{\theta_v} D_t\|$, which is determined as $\|\partial_{\theta_v^i} D_t\| = \|D_t^{-1}\text{diag}\left(\partial_{\theta_v^i} h_{j,t}\right)\|/2$, $i = 1,\cdots,3N$. The partial derivative of the j-th component above is zero when $i \neq j$. Otherwise, note that, by iterating the volatility process equation, we have

$$h_{j,t} = \frac{\varsigma_j}{1-\tau_j} + \kappa_j\left(\sum_{k\geq 1}\tau_j^{k-1}\epsilon_{j,t-k}^2\right),$$

$$\partial_{\varsigma_j} h_{j,t} = \frac{\varsigma_j}{1-\tau_j},\ \partial_{\kappa_j} h_{j,t} = \sum_{k\geq 1}\tau_j^{k-1}\epsilon_{j,t-k}^2,\ \text{and}\ \partial_{\tau_j} h_{j,t} = \frac{\varsigma_j}{(1-\tau_j)^2} + \sum_{k\geq 1}(k-1)\tau_j^{k-2}\epsilon_{j,t-k}^2.$$

We deduce there exists some constant C s.t.

$$\sup_{\theta:\|\theta_v-\theta_{0,v}<\alpha} \|\nabla_{\theta_v} D_t\|.\|\epsilon_t\|^2 \leq C\left(1+\sum_{k\geq 1}(k-1)\tau_j^{k-1}\epsilon_{j,t-k}^2\right)\|\epsilon_t\|^2 \text{ a.s.}$$

The latter r.h.s. belongs to L^1 because $E_{t-1}[\epsilon_{j,t}^2]=1$ for every j and t. Therefore,

$$\frac{1}{T}\sum_{t=1}^{T} \sup_{\theta:\|\theta_v-\theta_{0,v}<\alpha} \|\nabla_{\theta_v} D_t\|.\|\epsilon_t\|^2 = O_P(1),$$

proving (3.16) and then our lemma. □

3.2 Asymptotic Normality

We proved the consistency of $\hat{\theta}_{T,c}$. The consistency together with the central limit theorem are used to prove the asymptotic normality of $\hat{\theta}_T=(\hat{\theta}_{T,v},\hat{\theta}_{T,c})'$. To do so, several Taylor expansions are applied to the orthogonal conditions given by (1.7). Besides the assumptions defined for consistency, another set of hypothesis is required for the asymptotic normality.

Assumption 13. *$\theta_0 \in \mathring{\Theta}$ with $\mathring{\Theta}$ the interior of Θ.*

Assumption 14. *The innovations η_t have finite fourth order moments.*

Assumption 15. *Let the processes defined as*

$$\begin{aligned}
C_t &= (\nabla\Psi(Pc_{t-1}))^{-1}, \\
\tilde{C}_t &= (\nabla\Psi(Pc_{t-1}))^{-1}\left[\Xi\nabla\Psi(Pc_{t-1})+\Lambda\nabla\zeta(Pc_{t-1},D_{t-1},\epsilon_{t-1})\right], \\
E_t &= (\nabla\Psi(Pc_{t-1}))^{-1}\Psi(Pc_{t-1}), \\
G_t &= (\nabla\Psi(Pc_{t-1}))^{-1}\zeta(Pc_{t-1},D_{t-1},\epsilon_{t-1}),
\end{aligned}$$

The stochastic matrix process $(C_t,\tilde{C}_t,E_t,G_t)_t$ is stationary,

$$E\left[\|C_t\|^2+\|E_t\|^2+\|G_t\|^2\right]<+\infty, \text{ and } E\left[\|\sum_{k=1}^{\infty}\tilde{C}_{t-1}\tilde{C}_{t-2}\cdots\tilde{C}_{t-k}Z_{t-k-1}\|^2\right]<+\infty,$$

where the generic letter Z denotes C, E or G.

The next regularity conditions are classic and necessary to justify the existence of the asymptotic covariance in the next Theorem. They are assumed for convenience.

Under the price of additional technicalities, it is possible to establish some sufficient and more explicit conditions to satisfy the later ones.

Assumption 16. *$\nabla_{\theta_v\theta_c'} l_{2,t}(\epsilon_t;\theta_{0,v},\theta_{0,c})$, $\nabla_{\theta_c\theta_c'} l_{2,t}(\epsilon_t;\theta_{0,v},\theta_{0,c})$, $\nabla_{\theta_c\theta_c'}\psi(\theta_{0,v},\theta_{0,c})$ and $\nabla_{\theta_c\theta_v'}\psi(\theta_{0,v},\theta_{0,c})$ admit a finite first order moment.*

Assumption 17. *$\mathbb{E}\left[\nabla_{\theta_c\theta_c'} l_{2,t}(\epsilon_t;\theta_{0,v},\theta_{0,c})\right]$ is nonsingular.*

As expected, we need to assume that the initial values of the process are asymptotically irrelevant to evaluate score functions. The multiplication by $\sqrt{T}$ renders this task more difficult than in the proof of consistency. We have not tried to exhibit the equivalent of Lemma 3.2 to deal with this case.

Assumption 18. *$\sqrt{T}\|\Delta_T(\theta_{0,v})-\tilde{\Delta}_T(\theta_{0,v})\| = o_p(1)$ and $\sqrt{T}\|\Psi_T(\theta_{0,v},\theta_{0,c})-\tilde{\Psi}_T(\theta_{0,v},\theta_{0,c})\| = o_p(1)$.*

For some $\alpha > 0$, $\sup_{\theta_v:\|\theta_v-\theta_{0,v}\|<\alpha} \|\nabla_{\theta_v}\Delta_T(\theta_v) - \nabla_{\theta_v}\tilde{\Delta}_T(\theta_v)\| = o_p(1)$, *and*

$$\sup_{\theta:\|\theta-\theta_0\|<\alpha} \|\nabla_\theta \Psi_T(\theta) - \nabla_\theta \tilde{\Psi}_T(\theta)\| = o_p(1).$$

Theorem 3.3. *Assume Assumptions 4 and 8-18, then $\hat{\theta}_{T,v}$ and $\hat{\theta}_{T,c}$ are asymptotically normal, and*

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \xrightarrow{d} \mathcal{N}\left(0, J^{-1}IJ^{-1}\right),$$

where

$$J = \mathbb{E}_{\theta_0}\left[\begin{pmatrix} \nabla_{\theta_v\theta_v'} l_{1,t}(\epsilon_t;\theta_{0,v}) & 0 \\ \nabla_{\theta_v\theta_c'} l_{2,t}(\epsilon_t;\theta_{0,v},\theta_{0,c}) & \nabla_{\theta_c\theta_c'} l_{2,t}(\epsilon_t;\theta_{0,v},\theta_{0,c}) \end{pmatrix}\right],$$

$$I = \mathbb{E}_{\theta_0}\left[\begin{pmatrix} \nabla_{\theta_v} l_{1,t}(\epsilon_t;\theta_{0,v})\nabla_{\theta_v'} l_{1,t}(\epsilon_t;\theta_{0,v}) & \nabla_{\theta_v} l_{1,t}(\epsilon_t;\theta_{0,v})\nabla_{\theta_c'} l_{2,t}(\epsilon_t;\theta_{0,v},\theta_{0,c}) \\ \nabla_{\theta_c} l_{2,t}(\epsilon_t;\theta_{0,v},\theta_{0,c})\nabla_{\theta_v'} l_{1,t}(\epsilon_t;\theta_{0,v}) & \nabla_{\theta_c} l_{2,t}(\epsilon_t;\theta_{0,v},\theta_{0,c})\nabla_{\theta_c'} l_{2,t}(\epsilon_t;\theta_{0,v},\theta_{0,c}) \end{pmatrix}\right].$$

Remark about Assumption 16. This hypothesis for the two first quantities ensures the existence of the subblocks in J. The existence of the covariance of the scores in I is proved in Lemma (3.4) thanks to Assumption 15. Although the existence of the Hessians in J can be proved, it would require intense matrix computations and Lyapunov conditions on more complex coefficients than those given in Assumption 15. For the sake of clarity, we assume the existence of these Hessians.

Furthermore, assumption (16) for the last two quantities is used to prove the weak convergences in (i) of Lemma (3.5).

Remark about Assumption 17. This assumption ensures the invertibility of J, which requires the positive definiteness of $\mathbb{E}\left[\nabla_{\theta_v\theta_v'} l_{1,t}(\epsilon_t;\theta_{0,v})\right]$ and $\mathbb{E}\left[\nabla_{\theta_c\theta_c'} l_{2,t}(\epsilon_t;\theta_{0,v},\theta_{0,c})\right]$. Actually, the invertibility of the latter can be proved by contradiction based on a technical hypothesis, which is the family of vectors (vec($\partial_{\theta_c^i} R_t$) is linearly independent.

Lemma 3.4. *Suppose the assumptions of theorem (3.3) hold,*

(i) $\|\psi_t(\theta_{0,v},\theta_{0,c})\psi_t(\theta_{0,v},\theta_{0,c})'\|$, $\|\delta_t(\theta_{0,v})\psi_t(\theta_{0,v},\theta_{0,c})'\|$ *admit a finite first order moment.*

(ii) $\mathbb{V}_{as}\begin{pmatrix}\delta_t(\theta_{0,v})\\ \psi_t(\theta_{0,v},\theta_{0,c})\end{pmatrix} = I.$

Proof. *(i)* Note that the existence of $\mathbb{E}\left[\|\delta_t(\theta_{0,v})\delta_t(\theta_{0,v})'\|\right]$ and $\mathbb{E}\left[\|\nabla_{\theta_v}\delta_t(\theta_{0,v})\|\right]$ has been established by Francq and Zakoian (2004), as they are related to usual GARCH processes and Gaussian QMLE. This not require additional assumptions.

We denote by θ_v^i (resp. θ_c^i) the i-th component of the vector of volatility (resp. correlation) parameters. First we derive the score of the first step likelihood, which is in matrix form

$$l_{1,t}(\epsilon_t;\theta_v) = \log(|D_t^2(\theta_v)|) + \epsilon_t' D_t^{-2}(\theta_v)\epsilon_t.$$

For $i = 1,\cdots,3N$, after some matrix manipulations, this score function is given as

$$\begin{aligned}\delta_t^{(i)}(\theta_v) &= -\text{Trace}\left((\epsilon_t\epsilon_t' D_t^{-1} + D_t^{-1}\epsilon_t\epsilon_t')D_t^{-1}(\partial_{\theta_v^i} D_t)D_t^{-1}\right) + 2\text{Trace}\left(D_t^{-1}(\partial_{\theta_v^i} D_t)\right)\\ &= \text{Trace}\left(\left(I_N - D_t^{-1}\epsilon_t\epsilon_t' D_t^{-1}\right)\left(D_t^{-1}(\partial_{\theta_v^i} D_t) + (\partial_{\theta_v^i} D_t)D_t^{-1}\right)\right).\end{aligned}$$

Using $D_t^{-1}\epsilon_t = R_t^{1/2}\eta_t$, we obtain

$$\delta_t^{(i)}(\theta_v) = \text{Trace}\left(\left(I_N - R_t^{1/2}\eta_t\eta_t' R_t^{1/2}\right)\left(D_t^{-1}(\partial_{\theta_v^i} D_t) + (\partial_{\theta_v^i} D_t)D_t^{-1}\right)\right).$$

Choosing the spectral matrix norm, we have $\|R_t^{1/2}\| \leq \sqrt{\text{Tr}(R_t)} \leq \sqrt{N}$. Hence, $\delta_t^{(i)}(\theta_v)$ admits the upper bound

$$\begin{aligned}|\delta_t^{(i)}(\theta_v)| &\leq C_0.N\|I_N - R_t^{1/2}\eta_t\eta_t' R_t^{1/2}\|\|D_t^{-1}(\partial_{\theta_v^i} D_t) + (\partial_{\theta_v^i} D_t)D_t^{-1}\|\\ &\leq 2C_0N^2\left(1 + \|\eta_t\eta_t'\|\right).\|D_t^{-1}(\partial_{\theta_v^i} D_t)\|,\end{aligned}$$

for some constant C_0. The second step likelihood is defined as $l_{2,t}(\epsilon_t;\theta) = \log(|R_t(\theta)|) + u_t' R_t^{-1}(\theta) u_t$. For $i = 1, \cdots, 3N(N-1)/2$, its score function is

$$\psi_t^{(i)}(\theta_v, \theta_c) = \text{Trace}\left(\left(I_N - R_t^{-1/2}\eta_t\eta_t' R_t^{1/2}\right) R_t^{-1}(\partial_{\theta_c^i} R_t)\right).$$

This score can be upper bounded as

$$|\psi_t^{(i)}(\theta_v, \theta_c)| \leq C_0.N^2\left(1 + \lambda_{min}(R_t)^{-1}\|\eta_t\eta_t'\|\right)\|R_t^{-1}(\partial_{\theta_c^i} R_t)\|,$$

By the Cauchy-Schwartz inequality and Assumption 14, we have

$$\mathbb{E}\left[|\delta_t^{(i)}(\theta_v)\psi_t^{(j)}(\theta_v, \theta_c)|\right] \leq C_1\left\{1 + \mathbb{E}\left[\|R_t^{-1}(\partial_{\theta_c^j} R_t)\|^2\right]^{1/2}\right\}.\left\{1 + \mathbb{E}\left[\|D_t^{-1}(\partial_{\theta_v^i} D_t)\|^2\right]^{1/2}\right\},$$

for some constant $C_1 > 0$ and every $i = 1, \cdots, 3N$ and $j = 1, \cdots, 3N(N-1)/2$. Concerning the covariance of $\psi_t(\theta_v, \theta_c)$, we get similarly for every i, j

$$\mathbb{E}\left[|\psi_t^{(i)}(\theta)\psi_t^{(j)}(\theta)|\right] \leq C_2\left\{1 + \mathbb{E}\left[\|R_t^{-1}(\partial_{\theta_c^i} R_t)\|^2\right]^{1/2}\right\}.\left\{1 + \mathbb{E}\left[\|R_t^{-1}(\partial_{\theta_c^j} R_t)\|^2\right]^{1/2}\right\},$$

with $C_2 > 0$. Note that we have invoked the fact that the lower eigenvalue of R_t are bounded from above by a strictly positive constant, using Assumption 4. Therefore, it is sufficient to show that $\mathbb{E}\left[\|R_t^{-1}(\partial_{\theta_i} R_t)\|^2\right] < \infty$ and $\mathbb{E}\left[\|D_t^{-1}(\partial_{\theta_i} D_t)\|^2\right] < \infty$.

By the chain rule property, we have $\|\nabla_{\theta_c} R_t\| \leq \|\nabla F_{vine}(Pc_t).\nabla_{\theta_c} Pc_t\|$. But Assumption 4 implies that $F_{vine}(\cdot)$ is Lipschitz, i.e. its derivative is uniformly bounded. Now, setting $\theta_c^{(1)} = (\theta_c^i, i = 1, \ldots, N(N-1)/2) = \Omega$, we have

$$\begin{aligned}\nabla_{\theta_c^{(1)}} Pc_t &= (\nabla\Psi(Pc_{t-1}))^{-1} + (\nabla\Psi(Pc_{t-1}))^{-1}\left[\Xi\nabla\Psi(Pc_{t-1}) + \Lambda\nabla\zeta(Pc_{t-1}, D_{t-1}, \epsilon_{t-1})\right]\nabla_{\theta_c^{(1)}} Pc_{t-1} \\ &= C_{t-1} + \sum_{k=1}^{\infty}\tilde{C}_{t-1}\tilde{C}_{t-2}\cdots\tilde{C}_{t-k}C_{t-k-1},\end{aligned}$$

Furthermore, $\theta_c^{(2)} = (\theta_c^i, i = N(N-1)/2 + 1, \cdots, N(N-1)) = \text{diag}(\Xi)$, which is the vector stacking the diagonal element of Ξ, we obtain

$$\begin{aligned}\nabla_{\theta_c^{(2)}} Pc_t &= (\nabla\Psi(Pc_{t-1}))^{-1}\left[\Xi\nabla\Psi(Pc_{t-1}) + \Lambda\nabla\zeta(Pc_{t-1}, D_{t-1}, \epsilon_{t-1})\right]\nabla_{\theta_c^{(2)}} Pc_{t-1} \\ &+ (\nabla\Psi(Pc_{t-1}))^{-1}\Psi(Pc_{t-1}) \\ &= E_{t-1} + \sum_{k=1}^{\infty}\tilde{C}_{t-1}\tilde{C}_{t-2}\cdots\tilde{C}_{t-k}E_{t-k-1},\end{aligned}$$

Finally, $\theta_c^{(3)} = (\theta_c^i, i = N(N-1)+1, \cdots, 3N(N-1)/2) = \text{diag}(\Lambda)$, we get

$$\begin{aligned}
\nabla_{\theta_c^{(3)}} Pc_t &= (\nabla\Psi(Pc_{t-1}))^{-1} \left[\Xi\nabla\Psi(Pc_{t-1}) + \Lambda\nabla\zeta(Pc_{t-1}, D_{t-1}, \epsilon_{t-1})\right] \nabla_{\theta_c^{(3)}} Pc_{t-1} \\
&+ (\nabla\Psi(Pc_{t-1}))^{-1} \zeta(Pc_{t-1}, D_{t-1}, \epsilon_{t-1}) \\
&= G_{t-1} + \sum_{k=1}^{\infty} \tilde{C}_{t-1}\tilde{C}_{t-2}\cdots\tilde{C}_{t-k}G_{t-k-1},
\end{aligned}$$

Under Assumption 15, we deduce

$$E\left[\|R_t^{-1}(\partial_{\theta_i} R_t)\|^2\right] < \infty.$$

The existence of $\mathbb{E}\left[\|(\partial_{\theta_i} D_t) D_t^{-1}\|^2\right]$ was coming from the proof of Lemma 3.2. Hence, we have shown that, for $i = 1, \cdots, 3N$, and $j, k, l = 1, \cdots, 3N(N-1)/2$,

$$\mathbb{E}\left[|\delta_t^{(i)}(\theta_v)\psi_t^{(j)}(\theta_v, \theta_c)|\right] < \infty, \text{ and } \mathbb{E}\left[|\psi_t^{(k)}(\theta_v, \theta_c)\psi_t^{(l)}(\theta_v, \theta_c)|\right] < \infty,$$

proving the result.

(ii) Due to the orthogonal conditions, we have

$$\mathbb{V}_{as}\begin{pmatrix} \delta_t(\theta_{0,v}) \\ \psi_t(\theta_{0,v}, \theta_{0,c}) \end{pmatrix} = \begin{pmatrix} \mathbb{E}_{\theta_0}\left[\delta_t(\theta_{0,v})\delta_t(\epsilon; \theta_{0,v})\right] & \mathbb{E}_{\theta_0}\left[\delta_t(\theta_{0,v})\psi_t(\theta_{0,v}, \theta_{0,c})'\right] \\ \mathbb{E}_{\theta_0}\left[\psi_t(\theta_{0,v}, \theta_{0,c})\delta_t(\theta_{0,v})'\right] & \mathbb{E}_{\theta_0}\left[\psi_t(\theta_{0,v}, \theta_{0,c})\psi_t(\theta_{0,v}, \theta_{0,c})'\right] \end{pmatrix} = \Omega.$$

Now we focus on the Hessian matrix J, defined as

$$J = \begin{pmatrix} \mathbb{E}_{\theta_0}\left[\nabla_{\theta_v}\delta_t(\theta_{0,v})\right] & \mathbb{E}_{\theta_0}\left[\nabla_{\theta_c}\delta_t(\theta_{0,v})\right] \\ \mathbb{E}_{\theta_0}\left[\nabla_{\theta_v}\psi_t(\theta_{0,v}, \theta_{0,c})\right] & \mathbb{E}_{\theta_0}\left[\nabla_{\theta_c}\psi_t(\theta_{0,v}, \theta_{0,c})\right] \end{pmatrix} = \begin{pmatrix} J_{11} & J_{12} \\ J_{21} & J_{22} \end{pmatrix},$$

which is a lower triangular matrix as $\mathbb{E}_{\theta_0}\left[\nabla_{\theta_c}\delta_t(\theta_{0,v})\right] = 0$. Hence to prove J is positive definite, it is sufficient to prove that each block matrix on the diagonal of J is positive definite. By assumption, J_{22} is supposed positive definite. However, its proof can be lead based on a technical assumption stated in the following note.

□

Lemma 3.5. *Suppose the assumptions of theorem (3.3) hold. If $\bar{\theta}_T \to \theta_0$ in probability, then*

(i) $\nabla_{\theta_v}\Delta_T(\bar{\theta}_v) \xrightarrow[T\to\infty]{\mathbb{P}} J_{11}, \nabla_{\theta_v}\Psi_T(\bar{\theta}_v, \bar{\theta}_c) \xrightarrow[T\to\infty]{\mathbb{P}} J_{2,1}, \nabla_{\theta_c}\Psi_T(\bar{\theta}_v, \bar{\theta}_c) \xrightarrow[T\to\infty]{\mathbb{P}} J_{2,2}.$

(ii) $\sqrt{T}\begin{pmatrix} \Delta_T(\theta_{0,v}) \\ \Psi_T(\theta_{0,v}, \theta_{0,c}) \end{pmatrix} \xrightarrow{d} \mathcal{N}(0, I).$

Proof. *(i)* The first convergence corresponds to the sum of N scores of GARCH log likelihood. This was proved by Francq and Zakoian (2004).

We focus first on the last convergence. We apply a Taylor expansion of $\nabla_{\theta_c}\Psi(\hat{\theta}_T)$ around θ_0.

$$\begin{aligned}\nabla_{\theta_c}\Psi_T(\bar{\theta}_{T,v},\bar{\theta}_{T,c}) &= \nabla_{\theta_c}\Psi_T(\theta_{0,v},\theta_{0,c}) + \nabla_{\theta_c\theta_v'}\Psi_T(\tilde{\bar{\theta}}_{T,v},\bar{\theta}_{T,c})\left(\bar{\theta}_{T,v}-\theta_{0,v}\right)\\ &+ \nabla_{\theta_c\theta_c'}\Psi_T(\bar{\theta}_{T,v},\tilde{\bar{\theta}}_{T,c})\left(\bar{\theta}_{T,c}-\theta_{0,c}\right) + \circ_p(1),\end{aligned} \tag{3.18}$$

with $\tilde{\bar{\theta}}_T = x\theta_0 + (1-x)\bar{\theta}_T$, for $x \in]0,1[$. Furthermore, we apply the ergodic theorem to

$$\sup_{\theta:\|\theta-\theta_0\|<\alpha} \|\nabla_{\theta_c\theta_v'}\psi_t(\theta_v,\theta_c)\|, \qquad \sup_{\theta:\|\theta-\theta_0\|<\alpha} \|\nabla_{\theta_c\theta_c'}\psi_t(\epsilon_t;\theta_v,\theta_c)\|, \tag{3.19}$$

and by Theorem (3.1), we obtain $\tilde{\bar{\theta}}_T \underset{T\to\infty}{\longrightarrow} \theta_0$ a.s. Those two results imply

$$\begin{aligned}\lim_{T\to\infty}\sup \|\nabla_{\theta_c\theta_v'}\Psi_T(\theta_v,\theta_c)\| &\leq \lim_{T\to\infty}\sup \sum_{t=1}^{T} \sup_{\theta:\|\theta-\theta_0\|<\alpha} \|\nabla_{\theta_c\theta_v'}\psi_t(\theta_v,\theta_c)\|\\ &= \mathbb{E}\left[\sup_{\theta:\|\theta-\theta_0\|<\alpha} \|\nabla_{\theta_c\theta_v'}\psi_t(\theta_v,\theta_c)\|\right],\end{aligned} \tag{3.20}$$

and

$$\begin{aligned}\lim_{T\to\infty}\sup \|\nabla_{\theta_c\theta_c'}\Psi_T(\theta_v,\theta_c)\| &\leq \lim_{T\to\infty}\sup \sum_{t=1}^{T} \sup_{\theta:\|\theta-\theta_0\|<\alpha} \|\nabla_{\theta_c\theta_c'}\psi_t(\theta_v,\theta_c)\|\\ &= \mathbb{E}\left[\sup_{\theta:\|\theta-\theta_0\|<\alpha} \|\nabla_{\theta_c\theta_c'}\psi_t(\theta_v,\theta_c)\|\right].\end{aligned} \tag{3.21}$$

By Assumption 16, both expectations of (3.20) and (3.21) are finite. Besides, by Theorem (3.1), $\bar{\theta}_T \underset{T\to\infty}{\longrightarrow} \theta_0$ a.s., which implies that the two last terms of the r.h.s. of (3.18) converge to 0. Finally, the ergodic theorem applied to $\nabla_{\theta_c}\Psi_T(\theta_{0,v},\theta_{0,c})$ proves the last convergence of (i).

To prove the second convergence of (i), a Taylor expansion can be applied to $\nabla_{\theta_v}\psi_t(\theta_v,\theta_c)$. The same steps can be followed as previously: (ii) of Lemma (3.4), the strong consistency and the ergodic theorem.

(ii) We shall prove that the vector $(\delta_t(\theta_{0,v}),\psi_t(\theta_{0,v},\theta_{0,c}))'$ is a square integrable martingale difference to apply the central limit theorem of Billingsley.

For inference purposes, the correlation matrix is set to I_N in the first step estimation. The score with respect to the volatility components is given by

$$\delta_t^{(i)}(\theta_v) = \text{Trace}\left(\left(I_N - D_t^{-1}\epsilon_t\epsilon_t' D_t^{-1}\right)\left(D_t^{-1}(\partial_{\theta_v^i}D_t) + (\partial_{\theta_v^i}D_t)D_t^{-1}\right)\right).$$

Using $u_t = R_t^{1/2}\eta_t$ with $R_t = I_N$, $\mathbb{E}\left[\eta_t\eta_t'\right] = I_N$, and the $\mathcal{F}_{t-1}$ measurability of D_t, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\delta_t^{(i)}(\theta_v)|\mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\text{Trace}\left(\left(I_N - D_t^{-1}\epsilon_t\epsilon_t' D_t^{-1}\right)\left(D_t^{-1}(\partial_{\theta_v^i} D_t) + (\partial_{\theta_v^i} D_t)D_t^{-1}\right)\right)|\mathcal{F}_{t-1}\right] \\
&= 2\text{Trace}\left((\partial_{\theta_v^i} D_t)D_t^{-1})\right) - \text{Trace}\left(\mathbb{E}\left[u_t u_t'|\mathcal{F}_{t-1}\right](\partial_{\theta_v^i} D_t)D_t^{-1} + D_t^{-1}(\partial_{\theta_v^i} D_t))\right) \\
&= 2\text{Trace}\left((\partial_{\theta_v^i} D_t)D_t^{-1})\right) - 2\text{Trace}\left((\partial_{\theta_v^i} D_t)D_t^{-1})\right) \\
&= 0.
\end{aligned}
\tag{3.22}
$$

For the correlation components, for $i = 1, \cdots, 3N(N-1)/2$, the score is

$$
\psi_t^{(i)}(\theta_{0,v}, \theta_{0,c}) = \text{Trace}\left(\left(I_N - R_t^{-1}u_t u_t'\right) R_t^{-1}(\partial_{\theta_c^i} R_t)\right). \tag{3.23}
$$

Using $u_t = R_t^{1/2}\eta_t$, $\mathbb{E}\left[\eta_t\eta_t'\right] = I_N$, and the $\mathcal{F}_{t-1}$ measurability of R_t, we obtain

$$
\begin{aligned}
\mathbb{E}\left[\psi_t^{(i)}(\theta_{0,v}, \theta_{0,c})|\mathcal{F}_{t-1}\right] &= \mathbb{E}\left[\text{Trace}\left(\left(I_N - R_t^{-1}u_t u_t'\right) R_t^{-1}(\partial_{\theta_c^i} R_t)\right)|\mathcal{F}_{t-1}\right] \\
&= \text{Trace}\left(\left(I_N - R_t^{-1}\mathbb{E}\left[u_t u_t'|\mathcal{F}_{t-1}\right]\right) R_t^{-1}(\partial_{\theta_c^i} R_t)\right) \\
&= \text{Trace}\left(\left(I_N - R_t^{-1}R_t^{1/2}\mathbb{E}\left[\eta_t\eta_t'\right]R_t^{1/2}\right) R_t^{-1}(\partial_{\theta_c^i} R_t)\right) \\
&= \text{Trace}\left(\left(I_N - R_t^{-1}R_t\right) R_t^{-1}(\partial_{\theta_c^i} R_t)\right) \\
&= 0.
\end{aligned}
\tag{3.24}
$$

Consequently, $(\delta_t(\theta_{0,v}), \psi_t(\theta_{0,v}, \theta_{0,c}))'$ is a square integrable martingale difference.

The vector $(\delta_t(\theta_{0,v}), \psi_t(\theta_{0,v}, \theta_{0,c}))'$ is a function of ϵ_t together with elements, which are $\sigma(\epsilon_s, s < t)$ measurable. By assumption, the process ϵ_t is stationary. Consequently, by the central limit theorem of Billinsgley for stationary square integrable martingale difference, we have

$$
\sqrt{T}\begin{pmatrix} \Delta_T(\theta_{0,v}) \\ \Psi_T(\theta_{0,v}, \theta_{0,c}) \end{pmatrix} \xrightarrow{d} \mathcal{N}\left(0, \begin{pmatrix} \Omega_{11} & \Omega_{12} \\ \Omega_{21} & \Omega_{22} \end{pmatrix}\right). \tag{3.25}
$$

□

Proof. We now turn to the proof of Theorem (3.3). To do so, we shall apply a Taylor expansion around the first derivatives of the first and second step criteria. Expanding the first quantity in a Taylor series around $\theta_{0,v}$, we obtain

$$
0 = \Delta_T(\hat{\theta}_{T,v}) = \Delta_T(\theta_{0,v}) + \nabla_{\theta_v}\Delta_T(\bar{\theta}_{T,v})\left(\hat{\theta}_{T,v} - \theta_{0,v}\right),
$$

where $\|\bar{\theta}_{T,v} - \theta_{0,v}\| < \|\hat{\theta}_{T,v} - \theta_{0,v}\|$. Inverting this relationship and multiplying by

$\sqrt{T}$, we have

$$\sqrt{T}\left(\hat{\theta}_{T,v}-\theta_{0,v}\right)=\left(-\nabla_{\theta_v}\Delta_T(\bar{\theta}_{T,v})\right)^{-1}\sqrt{T}\;\Delta_T(\theta_{0,v}).$$

Since $\bar{\theta}_{T,v}\underset{T\to\infty}{\longrightarrow}\theta_{0,v}$, by the weak convergence of $\sqrt{T}\;\Delta_T(\theta_{0,v})$ and the Slutsky theorem, we obtain the result of Bollerslev and Wooldridge (1992)

$$\sqrt{T}\left(\hat{\theta}_{T,v}-\theta_{0,v}\right)\overset{d}{=}\mathcal{N}\left(0,A_0^{-1}B_0A_0^{-1}\right),\tag{3.26}$$

with $A_0=-\mathbb{E}\left[\nabla_{\theta_v}\delta_t(\theta_{0,v})\right]$ and $B_0=\mathbb{E}\left[\delta_t(\theta_{0,v})\delta_t(\theta_{0,v})'\right]$.

We now apply a Taylor expansion to the second step likelihood around $\theta_0=(\theta_{0,v},\theta_{0,c})$, such that $\|\bar{\theta}_{T,c}-\theta_{0,c}\|<\|\hat{\theta}_{T,c}-\theta_{0,c}\|$ and

$$0=\Psi_T(\hat{\theta}_{T,v},\hat{\theta}_{T,c})=\Psi_T(\theta_{0,v},\theta_{0,c})+\nabla_{\theta_v}\Psi_T(\bar{\theta}_{T,v},\bar{\theta}_{T,c})\left(\hat{\theta}_{T,v}-\theta_{0,v}\right)+\nabla_{\theta_c}\Psi_T(\bar{\theta}_{T,v},\bar{\theta}_{T,c})\left(\hat{\theta}_{T,c}-\theta_{0,c}\right).$$

Inverting this relationship and multiplying by $\sqrt{T}$, we obtain

$$\begin{aligned}\sqrt{T}\left(\hat{\theta}_{T,c}-\theta_{0,c}\right)&=\left(-\nabla_{\theta_c}\Psi_T(\bar{\theta}_{T,v},\bar{\theta}_{T,c})\right)^{-1}\nabla_{\theta_v}\Psi_T(\bar{\theta}_{T,v},\bar{\theta}_{T,c})\sqrt{T}\left(\hat{\theta}_{T,v}-\theta_{0,v}\right)\\&+\left(-\nabla_{\theta_c}\Psi_T(\bar{\theta}_{T,v},\bar{\theta}_{T,c})\right)^{-1}\sqrt{T}\;\Psi_T(\theta_{0,v},\theta_{0,c}).\end{aligned}$$

Using the expansion of the first step likelihood criterion, we obtain

$$\begin{aligned}\sqrt{T}\left(\hat{\theta}_{T,c}-\theta_{0,c}\right)&=\left(-\nabla_{\theta_c}\Psi_T(\bar{\theta}_{T,v},\bar{\theta}_{T,c})\right)^{-1}\nabla_{\theta_v}\Psi_T(\bar{\theta}_{T,v},\bar{\theta}_{T,c})\left(-\nabla_{\theta_v}\Delta_T(\bar{\theta}_{T,v})\right)^{-1}\sqrt{T}\;\Delta_T(\theta_{0,v})\\&+\left(-\nabla_{\theta_c}\Psi_T(\bar{\theta}_{T,v},\bar{\theta}_{T,c})\right)^{-1}\sqrt{T}\;\Psi_T(\theta_{0,v},\theta_{0,c}).\end{aligned}$$

Since $\bar{\theta}_{T,c}\underset{T\to\infty}{\longrightarrow}\theta_{0,c}$, by the weak convergence of $\sqrt{T}\;\Psi_T(\theta_{0,v},\theta_{0,c})$, by (ii) of Lemma (3.5) and the Slutsky theorem, we obtain

$$\sqrt{T}\begin{pmatrix}\Delta_T(\hat{\theta}_{T,v})-\Delta_T(\theta_{0,v})\\\Psi_T(\hat{\theta}_{T,v},\hat{\theta}_{T,c})-\Psi_T(\theta_{0,v},\theta_{0,c})\end{pmatrix}\overset{d}{\longrightarrow}\mathcal{N}\left(0,\begin{pmatrix}\Omega_{11}&\Omega_{12}\\\Omega_{21}&\Omega_{22}\end{pmatrix}\right).$$

Asymptotic normality is a consequence of the convergence in probability of the Hessian quantities, proved in (i) of Lemma (3.5), the convergence of the joint scores and the Slutsky theorem. As a by-product, simple calculations provide the asymptotic variances of $\hat{\theta}_{T,v}$ and $\hat{\theta}_{T,c}$: with obvious notations,

$$\mathbb{V}_{\text{as}}(\hat{\theta}_{T,v})=J_{11}^{-1}I_{11}J_{11}^{-1},$$

$$\mathbb{V}_{\text{as}}(\hat{\theta}_{T,c})=J_{22}^{-1}I_{22}J_{22}^{-1}-\Gamma I_{12}J_{22}^{-1}-J_{22}^{-1}I_{21}\Gamma'+\Gamma I_{11}\Gamma',\;\Gamma:=J_{22}^{-1}J_{21}J_{11}^{-1}.$$

To summarize the proof, we used Lemma (3.5) to prove that we can apply Taylor expansions to the likelihood functions with theoritical scores and Hessians as we only have the empirical counterparts. The main step for asymptotic normality is in (ii) in Lemma (3.5), which proves the asymptotic normality of the joint likelihood functions, the first step and second step. The weak convergence of the empirical Hessian to their theoritical counterparts is in step (i) of Lemma (3.4). The Slutsky theorem is finally used to prove the asymptotic normality of $\hat{\theta}_T$. Asymptotic normality also required the existence of the asymptotic variance covariance. This step is done in Lemma (3.4). □

Acknowledgement

The authors are grateful for very helpful discussions with C. Czado, C. Francq, O. Scaillet, J. Segers, and J.M. Zakoïan. They thank numerous seminar participants, particularly at the 11th World Congress of the Econometric Society, Computational and Financial Econometrics 2014 and 2015, the Bachelier seminar, the Crest-LFA seminar, the Oberwolfach conference about copulae 2015, etc. The authors thank the labex Ecodec for its support.

References

[1] Billingsley, P. (1995): *Probability and Measure.* New York: John Wiley and Sons, Inc.

[2] Bougerol, P. and Picard, N. (1992): *Stationarity of GARCH processes and of some nonnegative time series.* Journal of Econometrics 52, 1714-1729.

[3] Boussama, F., Fuchs, F. and Stelzer, R. (2011): *Stationarity and Geometric Ergodicity of BEKK Multivariate GARCH Models.* Stochastic Processes and their Applications 121, 2331-2360.

[4] Douc, R., Moulines, E. and Stoffer, D. (2014): *Nonlinear Time Series.* Chapman and Hall.

[5] Francq, C. and Zakoïan, J.M. (2004): *Maximum Likelihood Estimation of Pure GARCH and ARMA-GARCH Processes.* Bernoulli. 10(4), 605-637.

[6] Francq, C. and Zakoïan, J.M. (2010): *GARCH models.* Wiley.

[7] Higham, N.J. (2008): *Functions of Matrices.* SIAM.

[8] Kurowicka, D. and Cooke, R.M. (2006): *Completion Problem With Partial Correlation vines.* Linear Algebra and its Applications. 418, 188-200.

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

[10] Ling, S. and McAleer, M. (2003): *Asymptotic theory for a vector ARMA-GARCH model.* Econometric Theory 19, 280-310.

[11] Lütkepohl, H. (1996): *Handbook of Matrices.* Wiley.

[12] Tweedie, R.L. (1988):*Invariant measure for Markov chains with no irreductibility assumptions.* Journal of Applied Probability 25A, 275-285.

A Technical result: Assumption 10, Theorem (3.1)

Assumption 10 is proved in this section. It is probably the most difficult part as the nonlinear dynamic of R_t should be controlled. To prove Assumption 10, we need a technical assumption.

Assumption 19. *Ξ and Λ are diagonal matrices such that $\|\Xi\|_s < 1$, and $\mathbb{E}\left[\log\left(\|B_{t,m}(\chi,\epsilon)\|\right)\right] < 0$, where*

$$\mathbb{B}_{t-1,m}(\bar{\chi},\epsilon) = \begin{pmatrix} \frac{2}{\pi}\|\nabla_1\zeta_{t-1}\|\|\Lambda\| & \frac{2}{\pi}\|\nabla_1\zeta_{t-2}\|\|\Lambda\|\|\Xi\| & \cdots & \cdots & \frac{2}{\pi}\|\nabla_1\zeta_{t-m}\|\|\Lambda\|\|\Xi\|^{m-1} \\ 1 & 0 & \cdots & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 1 & 0 \end{pmatrix},$$

Above, $\zeta_t = \zeta\left(\chi_t,\eta_t\right)$ is the t-innovation of our partial correlation process, where $\chi_t = \left(\bar{P}c_t, \bar{D}_t\right)$ is a $\mathcal{F}_{t-1}$ measurable random vector, denoting by $\bar{P}c_t$ a random set of partial correlations that satisfies 4, and $\bar{D}_t$ is bounded a.e. Moreover, for $i = 1, 2$, $\nabla_i\zeta_t$ is the derivative of ζ_t with respect to its i-th component. Finally, $E[\|\epsilon_t\|^4] < \infty$.

Now Assumption 10 becomes

$$\sup_{\theta\in\Theta}|QL_{2,T}(\theta;\epsilon) - \widetilde{QL}_{2,T}(\theta;\epsilon)| \leq \frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|\log(|R_t|) - \log(|\tilde{R}_t|)| + \frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|u_t' R_t^{-1} u_t - \tilde{u}_t'\tilde{R}_t^{-1}\tilde{u}_t|. \quad \text{(A.1)}$$

We focus on the second sum, which can be written as

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|u_t' R_t^{-1} u_t - \tilde{u}_t'\tilde{R}_t^{-1}\tilde{u}_t| &= \frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|u_t'(R_t^{-1} - \tilde{R}_t^{-1})\tilde{u}_t + u_t' R_t^{-1}(u_t - \tilde{u}_t) + (u_t - \tilde{u}_t)'\tilde{R}_t^{-1}\tilde{u}_t| \\ &= \frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|\text{Trace}\left(u_t'(R_t^{-1} - \tilde{R}_t^{-1})\tilde{u}_t + u_t' R_t^{-1}(u_t - \tilde{u}_t) + (u_t - \tilde{u}_t)'\tilde{R}_t^{-1}\tilde{u}_t\right) \end{aligned}$$

By definition, $u_t = D_t^{-1}\epsilon_t$ and $\tilde{u}_t = \tilde{D}_t^{-1}\epsilon_t$. Thus, the previous quantity can be written as

$$\begin{aligned} &\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|\text{Tr}\left(\epsilon_t'\left[D_t^{-1}(R_t^{-1} - \tilde{R}_t^{-1})\tilde{D}_t^{-1} + D_t^{-1}R_t^{-1}(D_t^{-1} - \tilde{D}_t^{-1}) + (D_t^{-1} - \tilde{D}_t^{-1})\tilde{R}_t^{-1}\tilde{D}_t^{-1}\right]\epsilon_t\right)| \\ = &\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|\text{Tr}\left(\left[D_t^{-1}(R_t^{-1} - \tilde{R}_t^{-1})\tilde{D}_t^{-1} + D_t^{-1}R_t^{-1}(D_t^{-1} - \tilde{D}_t^{-1}) + (D_t^{-1} - \tilde{D}_t^{-1})\tilde{R}_t^{-1}\tilde{D}_t^{-1}\right]\epsilon_t\epsilon_t'\right)| \end{aligned}$$

We shall consider a multiplicative norm for matrices. To fix the ideas, this will be the spectral norm. Hence, we can bound the Trace operator as

$$\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}|\text{Tr}\left(\left[D_t^{-1}(R_t^{-1}-\tilde{R}_t^{-1})\tilde{D}_t^{-1}+D_t^{-1}R_t^{-1}(D_t^{-1}-\tilde{D}_t^{-1})+(D_t^{-1}-\tilde{D}_t^{-1})\tilde{R}_t^{-1}\tilde{D}_t^{-1}\right]\epsilon_t\epsilon_t'\right)$$

$$\leq \frac{N}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}(\|D_t^{-1}\|\|\tilde{R}_t^{-1}\|\|R_t-\tilde{R}_t\|\|R_t^{-1}\|\|\tilde{D}_t^{-1}\|+\|D_t^{-1}\|\|\tilde{D}_t^{-1}\|\|D_t-\tilde{D}_t\|\|D_t^{-1}\|(\|R_t^{-1}\|+\|\tilde{R}_t^{-1}\|))\|\epsilon_t\epsilon_t'\|$$

We denote

$$\begin{aligned}\mathbb{T}_t &= \|D_t^{-1}\|\|\tilde{R}_t^{-1}\|\|R_t-\tilde{R}_t\|\|R_t^{-1}\|\|\tilde{D}_t^{-1}\| \\ \mathbb{M}_t &= \|D_t^{-1}\|\|\tilde{D}_t^{-1}\|\|D_t-\tilde{D}_t\|\|D_t^{-1}\|\left(\|R_t^{-1}\|+\|\tilde{R}_t^{-1}\|\right)\end{aligned}$$

The main issue consists of controlling for $(R_t-\tilde{R}_t)$. We focus now on the quantity $\mathbb{T}_t$, and firstly on $\|R_t-\tilde{R}_t\|$.

$$\begin{aligned}R_t-\tilde{R}_t &= \text{vechof}(F_{vine}(Pc_t))-\text{vechof}(F_{vine}(\tilde{Pc}_t)), \\ &= \left[F_{vine}(Pc_t(i,j|L(i,j)))-F_{vine}(\tilde{Pc}_t(i,j|L(i,j)))\right]_{1\leq i,j\leq N}.\end{aligned}$$

Let $\epsilon>0$, and define the compact set $A_\epsilon=[-1+\epsilon,1-\epsilon]^{N(N-1)/2}$. The one-to-one mapping F_{vine} maps A_ϵ to $[-1+\tilde{\epsilon},1-\tilde{\epsilon}]^{N(N-1)/2}$, for some $\tilde{\epsilon}>0$. On A_ϵ, F_{vine} is C^1, hence ∇F_{vine} is bounded. Consequently, F_{vine} satisfies the Lipschitz condition: there exists $C>0$ s.t., for all x and $\tilde{x}\in A_\epsilon^2$, we have

$$\|F_{vine}(x)-F_{vine}(\tilde{x})\|_\infty\leq C\|x-\tilde{x}\|_\infty. \tag{A.2}$$

If we control the dynamics of these partial correlations, then we can ensure to generate trajectories within $[-1+\tilde{\epsilon},1-\tilde{\epsilon}]$. The stationary partial correlation processes are defined as

$$\Psi(Pc_t)=\Omega+\Xi\Psi(Pc_{t-1})+\Lambda\zeta_{t-1}. \tag{A.3}$$

When generating the partial correlation dynamics from arbitrarily fixed initial values, they are defined as

$$\Psi(\tilde{Pc}_t)=\Omega+\Xi\Psi(\tilde{Pc}_{t-1})+\Lambda\zeta_{t-1}. \tag{A.4}$$

In this process, the matrices are diagonal. Iterating (A.3), we get

$$\Psi(Pc_t) = \sum_{k=1}^{t} \Xi^{k-1}\Omega + \Xi^t \Psi(Pc_0) + \sum_{k=1}^{t} \Xi^{k-1}\Lambda\zeta_{t-k}, \tag{A.5}$$

where $\Psi(.)$ is applied to each component of the vector Pc_t and ζ_{t-k} is a function of Pc_{t-k}. The r.h.s. is an element of $\mathbb{R}^{N(N-1)/2}$. We recover Pc_t by inverting $\Psi(.)$ componentwise. (A.3) becomes

$$Pc_t = \Psi^{-1}(\sum_{k=1}^{t} \Xi^{k-1}\Omega + \Xi^t \Psi(Pc_0) + \sum_{k=1}^{t} \Xi^{k-1}\Lambda\zeta_{t-k}).$$

The trickiest part of this proof consists of controlling for the difference $Pc_t - \tilde{P}c_t$. The difficulty comes from the necessary transformation of ϵ_t, D_t and R_t to recover ζ_t. Now we have

$$\begin{aligned} Pc_t - \tilde{P}c_t &= \Psi^{-1}(\sum_{k=1}^{t} \Xi^{k-1}\Omega + \Xi^t \Psi(Pc_0) + \sum_{k=1}^{t} \Xi^{k-1}\Lambda\zeta_{t-k}) - \Psi^{-1}(\sum_{k=1}^{t} \Xi^{k-1}\Omega + \Xi^t \Psi(\tilde{P}c_0) \\ &+ \sum_{k=1}^{t} \Xi^{k-1}\Lambda\tilde{\zeta}_{t-k}) \\ &= \nabla\Psi^{-1}(X) \left[\Xi^t(\Psi(Pc_0) - \Psi(\tilde{P}c_0)) + \sum_{k=1}^{t} \Xi^{k-1}\Lambda(\zeta_{t-k} - \tilde{\zeta}_{t-k}) \right], \end{aligned}$$

for some matrix random X. The componentwise derivatives of Ψ^{-1} are the bounded functions $x \mapsto \dfrac{2}{\pi(1+x^2)}$. Hence $\|\nabla\Psi^{-1}\|_\infty \leq 2/\pi$ and we obtain

$$\|Pc_t - \tilde{P}c_t\| \leq \frac{2}{\pi}\|\Xi\|^t \|\Psi(Pc_0) - \Psi(\tilde{P}c_0)\| + \frac{2}{\pi}\|\Lambda\| \sum_{k=1}^{t} \|\Xi\|^{k-1} \|\zeta_{t-k} - \tilde{\zeta}_{t-k}\|,$$

where $\zeta_{t-k} = \zeta(\chi_{t-k}, \epsilon_{t-k})$, with $\chi_{t-k} = (Pc_{t-k}, D_{t-k})$. This gives the expansion

$$\zeta(\chi_{t-k}, \epsilon_{t-k}) - \zeta(\tilde{\chi}_{t-k}, \epsilon_{t-k}) = \nabla_1\zeta(\bar{\chi}_{t-k}, \epsilon_{t-k})(Pc_{t-k} - \tilde{P}c_{t-k}) + \nabla_2\zeta(\bar{\chi}_{t-k}, \epsilon_{t-k})(D_{t-k} - \tilde{D}_{t-k}),$$

where $\bar{\chi}_t$ is located between χ_t and $\tilde{\chi}_t$. Consequently, we deduce

$$\begin{aligned} \tfrac{\pi}{2}\|Pc_t - \tilde{P}c_t\| &\leq A_t + \tfrac{2}{\pi}\|\Lambda\| \sum_{k=1}^{t} \|\Xi\|^{k-1} \Big(\|\nabla_1\zeta(\bar{\chi}_{t-k}, \epsilon_{t-k})\| \|Pc_{t-k} - \tilde{P}c_{t-k}\| \\ &+ \|\nabla_2\zeta(\bar{\chi}_{t-k}, \epsilon_{t-k})\| \|D_{t-k} - \tilde{D}_{t-k}\| \Big), \end{aligned}$$

with $A_t = 2\|\Xi\|^t \|\Psi(Pc_0) - \Psi(\tilde{P}c_0)\|/\pi$. Denote $r_t = \|Pc_t - \tilde{P}c_t\|$ and $d_t = \|D_t - \tilde{D}_t\|$. Note that r_t is uniformly bounded, by a constant that depends on the considered norm. To simplify and wlog, this constant will be one here. We obtain

$$r_t \leq A_t + \frac{2}{\pi}\|\Lambda\| \sum_{k=1}^{t-1} \|\Xi\|^{k-1} \left(\|\nabla_1 \zeta(\bar{\chi}_{t-k}, \epsilon_{t-k})\| r_{t-k} + \|\nabla_2 \zeta(\bar{\chi}_{t-k}, \epsilon_{t-k})\| d_{t-k}\right). \quad \text{(A.6)}$$

Now we rewrite (A.6), for all $t \geq T$ and for some $m \leq t$ large enough that will be stated after, as

$$\vec{r}_{t,m} \leq \mathbb{C}_{t,m} + \mathbb{B}_{t-1,m}(\bar{\chi}, \epsilon)\vec{r}_{t-1,m}, \quad \text{(A.7)}$$

where $\mathbb{C}_{t,m} = \vec{A}_t + \vec{\mathcal{K}}_{t,m} + \vec{\mathcal{D}}_t$, and the vectors

$$\begin{aligned}
&\vec{r}_{t,m} = (r_t, r_{t-1}, \cdots, r_{t-m+1})', \ \vec{A}_t = (A_t, 0, \cdots, 0)', \ \vec{d}_{t,m} = (d_t, d_{t-1}, \cdots, d_{t-m+1})', \\
&\vec{\mathcal{K}}_{t,m} = \left(\tfrac{2}{\pi}\|\Lambda\| \sum_{k=m+1}^{t} \|\nabla_1 \zeta(\bar{\chi}_{t-k}, \epsilon_{t-k})\| \|\Xi\|^{k-1} r_{t-k}, 0, \cdots, 0\right)', \\
&\vec{\mathcal{D}}_t = \left(\tfrac{2}{\pi}\|\Lambda\| \sum_{k=1}^{t} \|\nabla_2 \zeta(\bar{\chi}_{t-k}, \epsilon_{t-k})\| \|\Xi\|^{k-1} d_{t-k}, 0, \cdots, 0\right)'.
\end{aligned} \quad \text{(A.8)}$$

These quantities are such that $\vec{r}_{t,m} \in \mathbb{R}^m$, $\vec{A}_t \in \mathbb{R}^m$, $\vec{\mathcal{K}}_{t-1,m} \in \mathbb{R}^m$, $\vec{\mathcal{D}}_t \in \mathbb{R}^m$.

We first focus on $\mathbb{C}_{t,m}$. For our matrix norm, we have

$$\|\mathbb{C}_{t,m}\| \leq \|\vec{A}_t\| + \|\vec{\mathcal{K}}_{t,m}\| + \|\vec{\mathcal{D}}_t\|. \quad \text{(A.9)}$$

Now iterating t in (A.7), let $0 < q < t$ fixed, we obtain

$$\vec{r}_{t,m} \leq \mathbb{C}_{t,m} + \sum_{k=1}^{q} \mathbb{B}_{t-1,m}(\bar{\chi}, \epsilon)\mathbb{B}_{t-2,m}(\bar{\chi}, \epsilon) \ldots \mathbb{B}_{t-k,m}(\bar{\chi}, \epsilon)\mathbb{C}_{t-k,m} + \mathbb{B}_{t-1,m}(\bar{\chi}, \epsilon) \cdots \mathbb{B}_{t-q-1,m}(\bar{\chi}, \epsilon)\vec{r}_{t-q-1,m}.$$

The sequence of matrices $\mathbb{B}_{t-k,m}(\bar{\chi}, \epsilon)$ is stochastic and each of them has a size depending on m. Under our assumptions, the series $\mathcal{B}_{t,m} := \sum_{k=1}^{+\infty} \prod_{j=1}^{k} \mathbb{B}_{t-j,m}(\bar{\chi}, \epsilon)$ is converging a.s. In particular, its main term tends to zero.

$$\begin{aligned}
\mathbb{P}\left(|\vec{r}_{t,m}| > \epsilon\right) &\leq \mathbb{P}(\|\mathbb{C}_{t,m}\| > \epsilon/3) + \mathbb{P}\Big(\prod_{j=1}^{q+1} \|\mathbb{B}_{t-j,m}(\bar{\chi}, \epsilon)\| > \epsilon/3\Big) \\
&+ \ \mathbb{P}\Big(\sum_{k=1}^{q} \prod_{j=1}^{k} \|\mathbb{B}_{t-j,m}(\bar{\chi}, \epsilon)\|.\|\mathbb{C}_{t-k,m}\| > \epsilon/3\Big) := T_1 + T_2 + T_3.
\end{aligned}$$

First, let us manage T_1, i.e. the $\mathbb{C}_{t,m}$ term. Since $\|\Psi(Pc_0) - \Psi(\tilde{P}c_0)\|$ is a fixed finite random variable and since $\|\Xi\| < 1$,

$$\mathbb{P}(\|A_t\| > \epsilon/9) < \epsilon,$$

for t sufficiently large (and independently of m and q). Moreover,

$$\mathbb{P}\left(\vec{\mathcal{K}}_{t,m} > \epsilon/9\right) \leq \mathbb{P}\left(\frac{2}{\pi}\|\Lambda\| \sum_{k=m+1}^{t} \|\nabla_1\zeta(\bar{\chi}_{t-k},\epsilon_{t-k})\|.\|\Xi\|^{k-1-m}.\|\Xi\|^m > \epsilon/9\right) \leq \epsilon,$$

for m sufficiently large and because the latter series converges a.s.

Denote by ρ the largest parameter among $\tau_1,\ldots,\tau_n$. By assumption, $\rho \in [0,1)$. Equation (4.6) in Francq and Zakoian (2004) provides $\sup_\theta \|D_t - \tilde{D}_t\| \leq K\rho^t$ a.s. Therefore,

$$\begin{aligned}\mathbb{P}\left(\|\vec{\mathcal{D}}_t\| > \epsilon/9\right) &\leq \mathbb{P}\left(\frac{2K}{\pi}\|\Lambda\|\sum_{k=1}^{t}\|\nabla_2\zeta(\bar{\chi}_{t-k},\epsilon_{t-k})\|\|\Xi\|^{k-1}\rho^{t-k} > \epsilon/9\right)\\ &\leq \mathbb{P}\left(\frac{2K\|\Lambda\|}{\pi t}\sum_{k=1}^{t}\|\nabla_2\zeta(\bar{\chi}_{t-k},\epsilon_{t-k})\|.t\max(\|\Xi\|,\rho)^{t-1} > \epsilon/9\right)\\ &\leq \epsilon\end{aligned}$$

for t sufficiently large, under our assumptions and the LLN. We deduce $T_1 \leq 3\epsilon$, for a well-chosen (and now fixed) m and for t sufficiently large.

Second, note that the main term of the series $\mathcal{B}_{t,m}$ tends to zero a.s. Therefore, $T_2 < \epsilon$ for the previous fixed m and q sufficiently large.

Third, it remains to deal with T_3. Actually, it is sufficient to use the same arguments as for T_1. Indeed,

$$\begin{aligned}&\mathbb{P}(\sum_{k=1}^{q}\prod_{j=1}^{k}\|\mathbb{B}_{t-j,m}(\bar{\chi},\epsilon)\|.\|\mathbb{C}_{t-k,m}\| > \epsilon/3) \leq \mathbb{P}(\sum_{k=1}^{q}\prod_{j=1}^{k}\|\mathbb{B}_{t-j,m}(\bar{\chi},\epsilon)\|.\|\vec{A}_{t-k,m}\| > \epsilon/9)\\ &+ \quad \mathbb{P}(\sum_{k=1}^{q}\prod_{j=1}^{k}\|\mathbb{B}_{t-j,m}(\bar{\chi},\epsilon)\|.\|\vec{\mathcal{K}}_{t-k,m}\| > \epsilon/9) + \mathbb{P}(\sum_{k=1}^{q}\prod_{j=1}^{k}\|\mathbb{B}_{t-j,m}(\bar{\chi},\epsilon)\|.\|\vec{\mathcal{D}}_{t-k,m}\| > \epsilon/9)\\ &:= \quad T_{31}+T_{32}+T_{33}.\end{aligned}$$

To be specific, due to the finiteness of $\mathcal{B}_{t,m}$,

$$T_{31} \leq \frac{2}{\pi}\mathbb{P}(\|\Psi(Pc_0)-\Psi(\tilde{P}c_0)\|.\|\Xi\|^{t-1}.\sum_{k=1}^{+\infty}\prod_{j=1}^{k}\|\mathbb{B}_{t-j,m}(\bar{\chi},\epsilon)\| > \epsilon/9),$$

that is less than ϵ for t sufficiently large (and a fixed m). The terms T_{32} and T_{33} are managed as above, because the multiplication by the (a.e. finite) random variable

$\mathcal{B}_{t,m}$ does not change the reasoning.

By grouping the all inequalities above and since the reasonings were uniform wrt θ, we get

$$\mathbb{P}\left(\sup_{\theta\in\Theta}|\vec{r}_{t,m}| > \epsilon\right) \leq 7\epsilon,$$

proving that $\sup_{\theta\in\Theta} r_t = o_P(1)$. Since it is bounded by one and due to the dominated convergence theorem, this convergence to zero is true in L^1 or L^2. This is true for $\|R_t - \tilde{R}_t\|$ too, because of (A.2): $\sup_{\theta\in\Theta}\|R_t - \tilde{R}_t\| = o_P(1)$ and $T^{-1}\sum_{t=1}^T \sup_{\theta\in\Theta}\|R_t - \tilde{R}_t\|$ tends to zero when $t\to\infty$.

We now focus on the precision matrix $R_t^{-1} := [\rho_t^{ij}]$. Obviously,

$$\rho_t^{ij} = (-1)^{i+j}\frac{det(R_t^{-(i,j)})}{det(R_t)},$$

where $R_t^{-(i,j)}$ is the covmatrix of R_t (the matrix deduced from R_t after having removed line i and column j). But note that Theorem 3.2 in Kurowicka and Cooke (2006) and Assumption 4 implies that there exists a constant a s.t. $det(R_t) > a > 0$ a.s. Since $det(R_t^{-(i,j)})$ is a finite sum of elements in $[-1,1]$, this term is bounded from above. Therefore, there exists a constant M_1 s.t.

$$\sup_{\theta\in\Theta}\|R_t^{-1}\| \leq M_1, \text{ a.s.}$$

The same argument holds for $\tilde{R}_t$: $\sup_{\theta\in\Theta}\|\tilde{R}_t^{-1}\| \leq M_2$.

Since $\|D_t^{-1}\|$, $\|\tilde{D}_t^{-1}\|$ and $\|R_t^{-1}\|$ are uniformly bounded from above, we deduce

$$\begin{aligned}\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^T\sup_{\theta\in\Theta}\mathbb{T}_t.\|\epsilon_t\epsilon_t'\| > \epsilon\right) &\leq \mathbb{P}\left(\frac{Cte}{T}\sum_{t=1}^T\sup_{\theta\in\Theta}\|Pc_t - \tilde{P}c_t\|.\|\epsilon_t\epsilon_t'\| > \epsilon\right)\\ &\leq \frac{Cte}{\epsilon}E\left[\sup_{\theta\in\Theta} r_t.\|\epsilon_t\epsilon_t'\|\right] \leq \frac{Cte}{\epsilon}E\left[\left(\sup_{\theta\in\Theta} r_t\right)^2\right]^{1/2}.E\left[\|\epsilon_t\epsilon_t'\|^2\right]^{1/2},\end{aligned}$$

that is less than ϵ for t sufficiently large.

The second term $\mathbb{M}_t$ can be bounded more straightforwardly. Using the stationarity assumption of the GARCH process, there exists $U > 0$, and $\rho \in]0,1[$ such that, a.s.,

$$\sup_{\theta\in\Theta}\sup_i\left|h_{i,t} - \tilde{h}_{i,t}\right| \leq U\rho^t.$$

Consequently, $\mathbb{M}_t$ can be bounded as

$$\sup_{\theta\in\Theta}\mathbb{M}_t = \sup_{\theta\in\Theta}\|D_t^{-1}\|\|\tilde{D}_t^{-1}\|\|D_t-\tilde{D}_t\|\|D_t^{-1}\|\left(\|R_t^{-1}\|+\|\tilde{R}_t^{-1}\|\right)\leq C\rho^t,\ \text{a.s}, \quad (A.10)$$

for some constant C. Then

$$\mathbb{P}\left(\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}\mathbb{M}_t\|\epsilon_t\epsilon_t'\|>\epsilon\right)\leq\mathbb{P}\left(\frac{C}{T}\sum_{t=1}^{T}\rho^t\|\epsilon_t\epsilon_t'\|>\epsilon\right)\leq\frac{C}{T\epsilon(1-\rho)}E\left[\|\epsilon_t\epsilon_t'\|\right]<\epsilon,$$

for t sufficiently large.

In other words, we have proved that

$$\frac{1}{T}\sum_{t=1}^{T}\sup_{\theta\in\Theta}(\mathbb{T}_t+\mathbb{M}_t).\|\epsilon_t\epsilon_t'\|=o_P(1).$$

For the first sum of (A.1) and considering the spectral norm, we have:

$$\begin{aligned}
\log(|R_t|)-\log(|\tilde{R}_t|) &= \log(|I_N+(R_t-\tilde{R}_t)\tilde{R}_t^{-1}|)\\
&\leq N\log(\|I_N+(R_t-\tilde{R}_t)\tilde{R}_t^{-1}\|)\\
&\leq N\log(\|I_N\|+\|(R_t-\tilde{R}_t)\tilde{R}_t^{-1}\|)\\
&\leq N\log(1+\|(R_t-\tilde{R}_t)\tilde{R}_t^{-1}\|)\\
&\leq N\|R_t-\tilde{R}_t\|\|\tilde{R}_t^{-1}\|.
\end{aligned}$$

By symmetry $\log(|\tilde{R}_t|)-\log(|R_t|)\leq N\|\tilde{R}_t-R_t\|\|R_t^{-1}\|$. Using the previous arguments, the first sum of (A.1) converges to 0 when $T\longrightarrow\infty$. We proved that

$$\sup_{\theta\in\Theta}|QL_{2,T}(\theta;\epsilon)-\widetilde{QL}_{2,T}(\theta;\epsilon)|=o_p(1). \quad (A.11)$$