High-dimensional penalized ARCH processes

Benjamin Poignard (Osaka University, MMDS)*

Jean-David Fermanian (Ensae-Crest)†

December 22, 2017

*E-mail address: poignard@sigmath.es.osaka-u.ac.jp

†5 av. Henry le Chatelier, 91120 Palaiseau, France. Tel.: +33170266715. E-mail address: jean-david.fermanian@ensae.fr (corresponding author)

Abstract

We introduce several families of multidimensional ARCH models, possibly with a very large number of parameters. The corresponding conditions of stationarity and of positive definiteness are studied. Through penalized OLS methods (sparse group-lasso), we consistently estimate such models. We evaluate the relevance of such strategies by simulation.

JEL classification: C13, C32, G17.

Keywords: Multivariate ARCH, Group-Lasso, stationarity, positive definiteness.

1 Introduction

Modeling the joint behavior of several financial assets has become a key challenge for academics and practitioners. Indeed, it is not easy to build a realistic model that is statistically relevant and consistent with some well-known stylized features of financial asset returns (fat tails, volatility clustering, autocorrelation of absolute returns, etc). In such discrete time multivariate framework, the usual key quantities are yielded by the covariance matrices of the current asset returns, given their past values. Indeed, an accurate estimation of covariance risk is crucial for risk management, asset pricing and portfolio management purpose.

In the literature, many specifications for discrete-time multivariate dynamic models have been proposed. Broadly speaking, most of them belong to the multivariate GARCH family or to the multivariate stochastic volatility family: see the surveys of Bauwens et al. (2006) and Asai et al. (2006), respectively. By specifying the dynamics of the first two conditional moments of the underlying distributions on one side, and the law of the innovations on the other side, such models are easy to simulate and to forecast one-step ahead. Nonetheless, in practical terms, a classical hurdle is related to the so-called "curse of dimensionality" as the specification of a general multivariate dynamic model often induces an explosion of the number of free parameters, inducing practical problems of inference and possibly overfitting.

Concerning N-dimensional GARCH models, the framework we adopt in the paper, inference is usually led by quasi-likelihood functions (see Francq and Zakoïan 2010, e.g.). The corresponding QML criteria are highly nonlinear - multivariate Gaussian or Student - with $O(N^2)$ free parameters and they necessitate fast solving optimization procedures. Therefore, strongly reduced versions of such multivariate models are most often considered as soon as N is larger than four or five, typically: the scalar BEKK of Engle and Kroner (1995), the scalar Dynamic Conditional Correlation (DCC) of Engle (2002) when modeling correlation processes, the Quadratic Flexible DCC of Billio and Caporin (2006), among others. However, it would be unrealistic to capture heterogenous patterns with such simplistic dynamic models, especially when N is "large". Indeed, with scalar models, the influence of past returns is similar for all components of the variance covariance matrix, a strong assumption.

Another approach is given by factor modeling, which aims at reducing the model complexity. Among others, Fan et al. (2008) emphasized the relevance of factor models for high-dimensional precision matrix estimation. However, this modeling ideas require the identification of corresponding factors. An "expert" approach is based on some priors regarding the leading underlying factors. Otherwise, latent unobserved factors induce particular estimation issues and their number is questionable.

The objective of this paper consists in modeling high-dimensional variance-covariance matrices within the multivariate GARCH framework, in a flexible way and breaking the curse of dimensionality. To do so, we introduce some extensions of the univariate ARCH model to multivariate ones, and we estimate such models through a convenient penalized ordinary least squares (OLS) procedure. Indeed, multivariate ARCH models admit a linear representation with respect to the parameters, contrary to GARCH ones. Note that any "invertible" GARCH process may be written as an infinite order ARCH model, under some conditions on its coefficients. Therefore, we argue that highly parameterized ARCH models (with numerous lags) should behave at least as well as more usual GARCH models, in terms of realism and flexibility. Nonetheless, for the purpose of parsimony and to avoid overfitting, we have to enforce the nullity of possibly numerous model coefficients. The OLS objective function is particularly adapted for regularization procedures and fast closed form-algorithms can be applied. A natural regularization procedure is given by the Sparse Group Lasso (SGL) of Simon et al. Tibshirani (2013), as it fosters sparsity at a group level and within a group, where the coefficients in the same group are associated to the same lag. We will consider an adaptive version of the SGL to satisfy the oracle property, which ensures the right identification of the underlying set of nonzero coefficients (Fan and Li 2001, Poignard 2016). In other words, we propose penalized OLS objective functions for a wide range of multivariate ARCH processes.

One of our main challenges is the non negativeness constraint for the generation of "true" conditional variance-covariance matrices. Indeed, in general, the model parameters must then satisfy highly nonlinear constraints. Then, the estimation problem is no longer convex and this prevents from using fast solving algorithms. Besides, the oracle property cannot be satisfied as it heavily relies on the convex property of the optimization criterion. To fix this issue, we propose several multivariate ARCH parameterizations that ensure non negativeness: the so-called homogeneous and heterogeneous ARCH models, and the Choleski-GARCH specification. To the best of our knowledge, the two former ones are new.

The paper is organized as follows. In Section 2, we describe the multivariate ARCH framework and our penalized ordinary least squares criteria. In Section 3, we introduce several highly parameterized ARCH-type models and discuss their stationarity property. In Section 4, we compare the performances of our penalized multivariate ARCH processes with other competitors by simulation.

2 The framework

2.1 High-dimensional ARCH-type specifications

We consider a N-dimensional vectorial stochastic process $(r_t)_{t=1,\cdots,T}$ and we denote by θ the vector of its model parameters. Typically, r_t is the vector of returns that is associated to a portfolio of financial assets. Typically, we decompose r_t as the sum of its conditional expected return and a residual:

$$r_t = \mu_t + \varepsilon_t, \;\; \varepsilon_t = H_t^{1/2}(\theta)\eta_t.$$

The expected return given the past is $\mu_t = \mathbb{E}[r_t|\mathcal{F}_{t-1}] := \mathbb{E}_{t-1}[r_t]$, where $\mathcal{F}_t$ denotes the market information until (and including) time t. To be short, $\mathcal{F}_t$ is the filtration induced the returns r_{t-k}, $k = 0, 1, 2, \ldots$. We set $H_t(\theta) = \mathrm{Var}(r_t|\mathcal{F}_{t-1}) := \mathrm{Var}_{t-1}(r_t) = \mathrm{Var}_{t-1}(\varepsilon_t)$ is a $N \times N$ that is a nonnegative matrix. The series (η_t) is supposed to be a strong white noise, i.e. a sequence of independent and identically distributed random variables s.t. $\mathbb{E}[\eta_t] = 0$ and $\mathrm{Var}(\eta_t) = I_N$, the identity matrix in $\mathbb{R}^N$. For convenience, we will denote $H_t(\theta) = H_t = [h_{k,l,t}]_{1\leq k,l\leq N}$.

The specification of the model above is complete when the law of η_t is specified and when the functional form of both μ_t and $H_t(\theta)$ are given. In this paper, we will focus on the centered dynamics (ε_t) after removing the first conditional moments. Now, these residuals will be considered as our observations (still denoted by ε_t). In practice, μ_t is estimated from the past returns and is $\mathcal{F}_{t-1}$-measurable. Therefore, keeping the same notations as above, the model we consider is actually $\varepsilon_t = H_t^{1/2}\eta_t$ for all t, where (η_t) is a $\mathcal{F}$-martingale difference: $\mathbb{E}[\eta_t|\mathcal{F}_{t-1}] = 0$ for all t. Moreover, we will assume that $\mathcal{F}_t = \sigma(\varepsilon_s, s \leq t)$.

The quantity of interest is H_t and we would like to directly specify its dynamics. A significant stream of the literature has been developed into this direction. A general formulation of H_t-dynamics has been proposed by Bollerslev et al. (1988): in their general VEC model, each element of H_t is a linear function of the lagged squared residuals, their cross-products and the components of lagged H_t matrices.

The most general formulation of a VEC(p,q) model is then

$$h_{i,j,t} = a_{i,j} + \sum_{k=1}^{q} \varepsilon'_{t-k} B_{ijk} \varepsilon_{t-k} + \sum_{l=1}^{p} C_{ij,l} \mathrm{vec}(H_{t-l}), \tag{2.1}$$

for every t and every indices i,j in $\{1,\ldots,N\}$. The model parameters are the unknown $N \times N$ matrices B_{ijk} and $C_{ij,l}$, $i,j \in \{1,\ldots,N\}$, $k = 1,\ldots,q$ and $l = 1,\ldots,p$. Moreover, $A := [a_{ij}]$ is a $N(N+1)/2$ unknown vector. Some tedious constraints have to be fulfilled to ensure that H_t is non negative in such a general parametrization. In this paper (and for some reasons that will appear hereafter), we will not consider the auto-regressive part in (2.1). Then, all matrices $C_{ij,l}$ are assumed to be zero and the model can now be rewritten as

$$H_t = A + \sum_{k=1}^{q} (I_N \otimes \varepsilon'_{t-k}) B_k (I_N \otimes \varepsilon_{t-k}), \tag{2.2}$$

where B_k is the $N^2 \times N^2$ block matrix given by $B_k := [B_{ijk}]_{1 \le i,j \le N}$ and $\otimes$ denotes the usual Kronecker product. In Gouriéroux (1997), it is stated that sufficient conditions for obtaining nonnegative covariance matrices H_t are the following ones:

(i) A and B_k, $k = 1,\ldots,q$, are symmetric, and

(ii) A and B_k, $k = 1,\ldots,q$, are non-negative.

Clearly, (i) can be imposed easily in the model specification and during the inference procedure, contrary to (ii). Indeed, in general, the latter condition imposes complex nonlinear constraints on the model parameters. Moreover, it is not realistic to estimate general non-negative matrices B, due to their sizes ($qN^2(N^2+1)/2$ unknown parameters!) and due to the tedious nonlinear constraints imposed by non-negativeness (particularly at the optimization stage). Therefore, we have to exhibit flexible but realistic sub-families of models given by (2.2). This will be done in Section 3.

Note that (2.2) can be written as a linear model

$$\varepsilon_t \varepsilon'_t = A + \sum_{k=1}^{q} (I_N \otimes \varepsilon'_{t-k}) B_k (I_N \otimes \varepsilon_{t-k}) + \zeta_t, \;\; \mathbb{E}[\zeta_t | \mathcal{F}_{t-1}] = 0. \tag{2.3}$$

To avoid redundancies, introduce the usual operator Vech(.) that transforms any $m \times m$ symmetric matrix M into the $m(m+1)/2$ vector of its component. Then, (2.3) is equivalent to

$$\mathrm{Vech}(\varepsilon_t \varepsilon_t') = \mathrm{Vech}(A) + \sum_{k=1}^{q} \mathrm{Vech}\left((I_N \otimes \varepsilon_{t-k}') B_k (I_N \otimes \varepsilon_{t-k})\right) + \mathrm{Vech}(\zeta_t).$$

This can be rewritten more explicitly: for every couple $(i,j) \in \{1,\dots,N\}^2$ such that $i \leq j$, we have

$$\varepsilon_{i,t}\varepsilon_{j,t} = a_{i,j} + \sum_{k=1}^{q} \sum_{r,s=1}^{N} b_{ijk,rs} \varepsilon_{r,t-k} \varepsilon_{s,t-k} + \zeta_{ij,t}, \ \mathbb{E}[\zeta_{ij,t} | \mathcal{F}_{t-1}] = 0, \tag{2.4}$$

where $B_{ijk} = [b_{ijk,rs}]_{1 \leq r,s \leq N}$. Note that the elements of the N^2-squared matrix B_k may be indexed by quadruplets (i,j,r,s), $1 \leq i,j,r,s \leq N$. The latter elements are related to the coefficients of B_k that define the dynamics of $\varepsilon_{i,t}\varepsilon_{j,t}$. Moreover, note that $B_{ijk} = B_{jik}$ and $\zeta_{ij,t} = \zeta_{ji,t}$ for every couple (i,j) and every k. Hereafter and if necessary, the couples of indices (i,j) and (r,s) will be sorted in the lexicographical order

$$(1,1),(1,2),\dots,(1,N),(2,1),(2,2),\dots,(N,N-1),(N,N),$$

even when we restrict ourselves to the couples (i,j) s.t. $i \leq j$.

The previous linear model will be estimated by a penalized least squares procedure. In terms of inference, this is a dramatic advantage w.r.t. the usual QML estimation procedure of GARCH models. Therefore, in practical terms, it will be easier to estimate ARCH-type models with a lot of assets and lags ($N >> 1$, $q >> 1$) than a GARCH model with the same N and $q = 1$.

2.2 A penalized empirical criterion

Contrary to GARCH-type dynamics that require the optimization of a nonlinear objective function (Gaussian- or Student-type likelihoods, typically), multivariate ARCH process have the advantage of allowing direct estimation by ordinary least squares. Assume that the true model is (2.4), with the true indices p_0 and q_0. A regularization procedure with q larger than q_0 would likely set the parameters $b_{ijk,rs}$ to zero when $k > q_0$. Moreover, note that, if the true model is a GARCH process, then it can be rewritten as in (2.4) with $q = \infty$ (if a convenient bock-companion matrix of the autoregressive parameters is invertible, strictly speaking). In such a case, the model (2.4) may produce relevant approximations of usual GARCH processes. Since q_0 is unknown, these arguments call for choosing a "sufficiently large" q ex ante.

For the sake of parsimony, the estimated parameters need to be constrained to avoid overfitting. The OLS objective function is particularly adapted to penalized procedures. The asymptotic properties of the associated estimators can be found in Fan and Li (2001), for instance. Such a regularization procedure aims at identifying the relevant subset of parameters, to describe the instantaneous covariances. A priori, the parameter θ belongs to a bigger set formed by some (possibly numerous) lagged variables. Both estimation and variable selection will be performed through regularization.

Now, let us specify such a well-suited procedure to be applied to some high-dimensional ARCH models. Our "non-penalized" least squares objective function will be

$$\begin{cases} \mathbb{G}_T l(\theta) & = & \frac{1}{T}\sum_{t=1}^{T} l(\varepsilon_t;\theta), \\ l(\varepsilon_t;\theta) & = & \|\text{Vech}(\varepsilon_t\varepsilon_t') - \Psi(\underline{\varepsilon_{t-1}})\theta\|_2^2, \end{cases} \tag{2.5}$$

where $\Psi(\varepsilon_{t-1})$ is a $\mathcal{F}_{t-1}$-measurable random matrix, whose particular analytic form depends on the model specification. For instance, for the process (2.4) and without any additional constraint on the parameters, the parameter vector can be decomposed as $\theta = (\theta^{(ij)}, 1 \leq i \leq j \leq N)$, such that the ij-th sub-vector is

$$\theta^{(ij)} := (a_{ij}, \theta^{(ij1)}, \dots, \theta^{(ijq)}),$$

$$\theta^{(ijk)} := (b_{ijk,11}, 2b_{ijk,12}, \cdots, 2b_{ijk,1N}, b_{ijk,22}, 2b_{ijk,23}, \dots, 2b_{ijk,(N-1)N}, b_{ijk,NN})'.$$

This means that the number of unknown parameters is $d(1+qd)$, with $d = N(N+1)/2$. Then, in such a case, $\Psi(\underline{\varepsilon_t})$ is the $d \times d(1+qd)$ matrix

$$\Psi(\underline{\varepsilon_t}) = \begin{pmatrix} \psi(\underline{\varepsilon_t}) & 0_{1+qd} & 0_{1+qd} & 0_{1+qd} & \cdots & 0_{1+qd} \\ 0_{1+qd} & \psi(\underline{\varepsilon_t}) & 0_{1+qd} & 0_{1+qd} & \cdots & 0_{1+qd} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0_{1+qd} & 0_{1+qd} & 0_{1+qd} & \cdots & 0_{1+qd} & \psi(\underline{\varepsilon_t}) \end{pmatrix},$$

where 0_{1+qd} is a $1+qd$-row vector of zeros and

$$\psi(\underline{\varepsilon_t}) = (1, \text{Vech}(\varepsilon_{t-1}\varepsilon'_{t-1})', \ldots, \text{Vech}(\varepsilon_{t-q}\varepsilon'_{t-q})').$$

Note that the latter criterion has most often to be rewritten when some constraints on the model parameters are taken into account. Indeed, in such a case, the number of free parameters is typically reduced, and/or some parameters are shared by several univariate linear equations of the type (2.4). See, for instance, the so-called "homogeneous model" in Subsection 3.2.

Moreover, in a lot of situations, it is likely that the most recent observations should have a higher level effect on the current covariance matrix than older observations. Think of a usual univariate GARCH$(1,1)$ process and its ARCH(∞) rewriting, for instance. In this setting it is natural to assume that the model parameters $b_{ijk,rs}$ decay with k, i.e. as we move farther away from the current observation. We could consider a procedure that would impose inequality constraints among the coefficients to recover such ordering effects. Following the same intuition, Tibshirani and Suo (2016) proposed an order-constrained version of the Lasso. Such additional constraints can easily be added into our framework. To lighten the presentation, we have not explicitly considered them hereafter. At least, we only assume that all the coefficients are zero from a certain rank $k \leq q$ on.

Now, let us penalize the previous OLS criterion to foster parsimony. The intuition is as follows: after having specified a large number of lags q a priori, assume that only a subset of potential lagged variances and covariances produce a statistically significant effect on the current covariances (the sparsity assumption). A penalization procedure enables to recover this unknown subset by enforcing some estimated coefficients to zero. Among a lot of competitors (Lasso, SCAD, elastic-net, etc), the Sparse Group Lasso seems to be the most relevant regularizer as it fosters sparsity both at a group level and within a group. Intuitively, the natural groups should be all the parameters that are associated to a given lagged vector ε_{t-k} (i.e. all quantities $b_{ijk,rs}$ for every quadruplet (i,j,r,s)), but other choices are possible, obviously.

To fix the ideas, potentially every component of θ belongs to some subvector $\theta^{(k)}$, $k=1,\ldots,m$, whose size is denoted by $\boldsymbol{c}_k$. In other words, the concatenation of all $\theta^{(k)}$ provides θ (or a subset of θ), after a rearrangement of its components. In our "core" example, $m=q$ and we concatenate into $\theta^{(k)}$ all coefficients $b_{ijk,rs}$ for every (i,j,r,s). Even possible, we will not penalize the coefficients $a_{i,j}$ because we will propose to estimate them in a preliminary stage through a targeting procedure (see Subsection 2.3).

Then, our statistical problem consists in minimizing over some finite-dimensional parameter space Θ a penalized criterion of the form

$$\hat{\theta} = \arg\min_{\theta\in\Theta} \{\mathbb{G}_T\varphi(\theta)\}, \tag{2.6}$$

where $\mathbb{G}_T l(\theta) + \boldsymbol{p}_1(\lambda_T,\tilde{\theta},\theta) + \boldsymbol{p}_2(\gamma_T,\tilde{\theta},\theta)$. Both penalties are specified as

$$\begin{cases} \boldsymbol{p}_1 : \mathbb{R}_+ \times \Theta \times \Theta \to \mathbb{R}_+, & \boldsymbol{p}_2 : \mathbb{R}_+ \times \Theta \times \Theta \to \mathbb{R}_+, \\ (\lambda_T,\tilde{\theta},\theta) \mapsto \boldsymbol{p}_1(\lambda_T,\tilde{\theta},\theta) = \dfrac{\lambda_T}{T}\sum\limits_{k=1}^{m}\sum\limits_{i=1}^{\boldsymbol{c}_k}\alpha_{T,i}^{(k)}|\theta_i^{(k)}|, & (\gamma_T,\tilde{\theta},\theta) \mapsto \boldsymbol{p}_2(\gamma_T,\tilde{\theta},\theta) = \dfrac{\gamma_T}{T}\sum\limits_{l=1}^{m}\xi_{T,l}\|\theta^{(l)}\|_2, \end{cases}$$

with $\alpha_{T,i}^{(k)} = |\tilde{\theta}_i^{(k)}|^{-\eta}$ and $\xi_{T,l} = \|\tilde{\theta}^{(l)}\|_2^{-\mu}$, where $\eta>0, \mu>0$, and $\tilde{\theta}$ is a first step estimator of θ, which is supposed to be $\sqrt{T}$-consistent. For instance, $\tilde{\theta}$ can be an unpenalized OLS estimator. Its $\sqrt{T}$-consistency is necessary to satisfy the oracle property. The tuning parameters λ_T and γ_T typically tend to zero when $T\to\infty$

(see Poignard 2016).

This program reduces to the classic OLS estimator when there is no penalization. The proposed penalization framework includes the usual Lasso criterion when $\gamma_T = 0$, the Group Lasso when $\lambda_T = 0$ and the Sparse Group Lasso when λ_T and γ_T are non zero.

Obtaining the non negativeness of the conditional covariance matrices induced by (2.4) is the main technical challenge in practice. To ensure this constraint, the parameters in (2.4) must satisfy eigenvalue-type constraints such that Θ will not be convex. This is a drawback from both an empirical and theoretical point of views: empirically, it hampers fast solving algorithms; theoretically, the non-convexity prevents the Sparse Group Lasso estimator from satisfying the oracle property of Fan and Li (2001). Thus, in Section 3, we propose parameterizations that allow for generating non negative matrices while remaining flexible and linear with respect to the parameters. This would discard processes that require a normalization step or non convex constraint sets for the parameters.

2.3 Evaluation of A

As a digression, let us focus on a covariance targeting procedure for the estimation of A. Although this parameter could be estimated with B simultaneously, the covariance targeting step fosters dimension reduction as it splits the problem. This will allow to satisfy the non-negativeness of the estimated matrix A more easily. To do so, note that taking the unconditional expectation of (2.4), we have

$$\mathbb{E}[\varepsilon_{i,t}\varepsilon_{j,t}] = a_{i,j} + \sum_{k=1}^{q} \sum_{r,s=1}^{N} b_{ijk,rs} \mathbb{E}[\varepsilon_{r,t-k}\varepsilon_{s,t-k}],$$

for every couple (i, j). If the coefficients $b_{ijk,rs}$ were known, and assuming we have estimated consistently $\mathbb{E}[\varepsilon_{i,t}\varepsilon_{j,t}]$ by $\widehat{\text{cov}}_{i,j}$, then the coefficients $a_{i,j}$ could be estimated as

$$\hat{a}_{i,j} = \widehat{\text{cov}}_{i,j} - \sum_{k=1}^{q} \sum_{r,s=1}^{N} b_{ijk,rs} \widehat{\text{cov}}_{r,s}.$$

When T is large and assuming the model is well specified, $\hat{a}_{i,j}$ will converge towards $a_{i,j}$ and we would observe that the estimated matrix $\hat{A} := [\hat{a}_{i,j}]$ is definite positive if this is the case for A. Nonetheless, at finite distance, it is likely the latter condition will not be satisfied. Fortunately, our OLS estimation procedure does not require per se that we manipulate nonnegative matrices A and B. This is required only for prediction and likelihood-based methods. Therefore, to estimate (2.2) (and then (2.4)), we propose to replace $a_{i,j}$ by $\hat{a}_{i,j}$, and the model is then parameterized by B only. Once B is estimated by $\hat{B}$, the matrix A will be approximated by $\tilde{A}$ whose components are

$$\tilde{a}_{i,j} = \widehat{\text{cov}}_{i,j} - \sum_{k=1}^{q} \sum_{r,s=1}^{N} \hat{b}_{ijk,rs} \widehat{\text{cov}}_{r,s}.$$

Afterwards, a projection of $\tilde{A}$ on the cone of nonnegative matrices would provide the final estimate of A.

As an alternative strategy, we can invoke a parametrization of A in the cone of nonnegative matrices directly. The natural basis would be provided by the spectral decomposition of $\mathbb{E}[\varepsilon_t \varepsilon_t']$ (or its empirical approximation $[\widehat{\text{cov}}_{i,j}]$ instead). Indeed, there exists an orthonormal family $(\boldsymbol{v}_1, \ldots, \boldsymbol{v}_N)$ in $\mathbb{R}^N$ s.t.

$$\mathbb{E}[\varepsilon_t \varepsilon_t'] \simeq [\widehat{\text{cov}}_{i,j}]_{1 \leq i,j \leq N} = \sum_{l=1}^{N} \nu_l \boldsymbol{v}_l \boldsymbol{v}_l',$$

where $(\nu_1, \ldots, \nu_N)$ is the associated spectrum, $\nu_1 \geq \nu_2 \geq \cdots \geq \nu_N \geq 0$. Then, we could assume that there exist nonnegative real numbers π_l, $l = 1, \ldots, N$ s.t. $A = \sum_{l=1}^{N} \pi_l \boldsymbol{v}_l \boldsymbol{v}_l'$. Then, we have replaced the $N(N+1)/2$ unknown coefficients of A by only N parameters $(\pi_1, \ldots, \pi_N)$. And such a matrix A will be nonnegative by construction.

In our applications, we have used the first targeting method for A. From now on, we focus on the evaluation of B-type matrices in (2.4).

3 Our ARCH-type specifications

In this section, we propose several ARCH-type parameterizations of (2.2) to ensure the non negativeness of H_t. Remind that our main objective is to obtain linear processes whose parameters possibly satisfy linear constraints. These are sufficient conditions to obtain a convex objective function on a convex parameter set. First, we propose a constraint free multivariate ARCH dynamics (the B-parameters are unconstrained) and the corresponding (H_t) process is projected onto the space of nonnegative matrices. The second model is called "homogeneous" and is relevant for random vectors with positively correlated components. Then, we propose a "heterogenous" parametrization that it is adapted to random vectors with discordant patterns. Finally, a model based on Choleski decompositions is discussed.

3.1 Constraint free and matrix projection

This "brute-force" approach consists in projecting a matrix process, which may not be necessarily non negative, onto $\mathcal{M}^+_{N\times N}(\mathbb{R})$, the cone of non negative matrices. This method allows flexibility because one can independently specify and estimate the processes that are associated to each component of $\text{vec}(\varepsilon_t\varepsilon_t')$. We rewrite the general dynamics given by (2.4) for each component of the $\varepsilon_t\varepsilon_t'$ matrix as

$$\varepsilon_{i,t}\varepsilon_{j,t} = a_{i,j}+\sum_{k=1}^{q}\sum_{r=1}^{N} b_{ijk,rr}\varepsilon^2_{r,t-k}+\sum_{k=1}^{q}\sum_{r,s=1,r<s}^{N} 2b_{ijk,rs}\varepsilon_{r,t-k}\varepsilon_{s,t-k}+\zeta_{i,j,t},\ \mathbb{E}[\zeta_{i,j,t}|\mathcal{F}_{t-1}]=0, \tag{3.1}$$

if $i \leq j$. Through inference by OLS, the symmetric matrices A and B are not necessarily non negative. Nonetheless, these matrices can be approximated by nonnegative ones. Here is a cost to be paid: eventually, we no longer satisfy (3.1) strictly speaking, to generate true conditional covariance matrices (H_t).

To this goal, consider the singular value decomposition of a symmetric matrix M as $M = P'\text{diag}(\lambda_1, \cdots, \lambda_N)P$, where P is an orthogonal matrix composed with N eigenvectors. We define two projections $f_k : \mathcal{M}_{N\times N}(\mathbb{R}) \to \mathcal{M}^+_{N\times N}(\mathbb{R})$, $k = 1, 2$. A first projection is $f_1(M) = P'\text{diag}(\lambda^+_1, \cdots, \lambda^+_N)P$, with λ^+_k the positive part of

λ_k. A second projection is $f_2(M) = (M + \lambda^-_{\min} I_d)/(1 + \lambda^-_{\min})$, with $\lambda^-_{\min}$ the negative part of the minimum eigenvalue of M. The eigenvectors remain the same as for M in both cases. Note that we can even impose the positive definiteness (no zero eigenvalue) of the projected matrices by adding cI_N to $f_k(M)$, for some arbitrarily small positive number c.

The first stage estimated matrix is denoted by $\tilde{H}_t = [\tilde{h}_{ij,t}]$, whose components are given by

$$\tilde{h}_{ij,t} = \hat{a}_{i,j} + \sum_{k=1}^{q} \sum_{r=1}^{N} \hat{b}_{ijk,rr} \varepsilon^2_{r,t-k} + \sum_{k=1}^{q} \sum_{r,s=1,r<s}^{N} 2\hat{b}_{ijk,rs} \varepsilon_{r,t-k} \varepsilon_{s,t-k},$$

for any couple (i,j). For any projection method $k \in \{1,2\}$, the final estimated covariance matrix of ε_t given $\mathcal{F}_{t-1}$ would be $H_t = f_k(\tilde{H}_t)$.

This method allows for an equation-by-equation estimation procedure, where each equation corresponds to a couple (i,j). This feature is particularly adapted for high-dimensional regression settings. Such dynamics are linear with respect to the parameters so that the estimation can be carried out by the ordinary least squares objective function or by penalized OLS.

3.2 The homogeneous portfolio model

Here, we particularize the general ARCH model (2.4). We will need some matrix notations:

- For any subset J of indices in $I := \{1, \ldots, m\}$, the m-column vector $e_{m,J}$ of zeros and ones is defined by $e_{m,J} := [\mathbf{1}(i \in J)]_{1 \leq i \leq m}$. When its size is obvious, it is written e_J simply. Moreover, $e_{m,I} = e_m$ is the m-vector of ones.

- For any vector $\boldsymbol{x} \in \mathbb{R}^m$, $D(\boldsymbol{x})$ denotes the $m \times m$ diagonal matrix given by $D(\boldsymbol{x}) = [\mathbf{1}(i = j) x_i]_{1 \leq i,j \leq m}$.

Set $\mathcal{J} = \{1, N+2, 2N+3, \ldots, (N-2)N+N-1, (N-1)N+N\}$, a subset of $\{1, \ldots, N^2\}$. Let us consider the parametric family $\mathcal{B}$ of matrices given by

$$\mathcal{B} = \{M \in \mathcal{M}_{N^2 \times N^2}(\mathbb{R}) \,|\, M = \alpha e_{N^2} e'_{N^2} + \beta e_{\mathcal{J}} e'_{\mathcal{J}} + \gamma D(e_{\mathcal{J}}), (\alpha, \beta, \gamma) \in [0,1]^3\}.$$

Clearly, all matrices in $\mathcal{B}$ are non-negative. By assumption, we will choose our matrices B_k, $k = 1, \ldots, q$, inside $\mathcal{B}$. More explicitly, in the homogenous ARCH model, we have for every indices i, j and time t

$$\varepsilon_{it}\varepsilon_{jt} = a_{ij} + \sum_{k=1}^{q} \left((\alpha_k + \beta_k + \gamma_k \mathbf{1}(i=j)) \varepsilon_{i,t-k}\varepsilon_{j,t-k} + \alpha_k \sum_{(r,s) \neq (i,j)} \varepsilon_{r,t-k}\varepsilon_{s,t-k} \right) + \zeta_{ij,t},$$

where $\zeta_{ij,t} = \varepsilon_{it}\varepsilon_{jt} - h_{ij,t}$. Note that the matrix $e_{\mathcal{J}} e'_{\mathcal{J}}$ can be rewritten as a block-matrix $[E_{ij}]_{1 \leq i,j \leq N}$, where $E_{ij} = [\mathbf{1}((i,j) = (r,s))]_{1 \leq r,s \leq 1}$.

This model specification tries to simultaneously capture three effects on the dynamics of $\varepsilon_{i,t}\varepsilon_{j,t}$:

(i) a uniform effect of all past cross-product among the components of $\varepsilon_t \varepsilon'_t$ through the α_k coefficients;

(ii) a more important bump caused by the past values of $\varepsilon_{i,t}\varepsilon_{j,t}$ on itself through β_k;

(iii) an additional bump when variances are managed (ie when $i = j$) through the parameters γ_k.

As for the estimation step, the underlying unknown parameter corresponds to

$$\theta = (\alpha_1, \ldots, \alpha_q, \beta_1, \ldots, \beta_q, \gamma_1, \ldots, \gamma_q,),$$

when the constant $a_{i,j}$ has been removed as explained in Subsection 2.3. In this case, we can apply the penalized OLS procedure, as detailed in Subsection 2.2. The matrix $\Psi(\underline{\varepsilon_{t-1}})$ of regressors is then

$$\Psi(\underline{\varepsilon_{t-1}}) = \begin{pmatrix} s_{t-1} & \dots & s_{t-q} & \vec{\varepsilon}_{11,t,q} & \vec{\varepsilon}_{11,t,q} \\ s_{t-1} & \dots & s_{t-q} & \vec{\varepsilon}_{12,t,q} & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ s_{t-1} & \dots & s_{t-q} & \vec{\varepsilon}_{NN,t,q} & \vec{\varepsilon}_{NN,t,q} \end{pmatrix},$$

where $s_{t-k} := \sum_{r,s=1}^{N} \varepsilon_{r,t-k}\varepsilon_{s,t-k}$, for $k = 1, \dots, q$ and $\vec{\varepsilon}_{ij,t,q} := (\varepsilon_{i,t-1}\varepsilon_{j,t-1}, \cdots, \varepsilon_{i,t-q}\varepsilon_{j,t-q})$. Note that the size of $\Psi(\underline{\varepsilon_{t-1}})$ is here $N(N+1)/2 \times 3q$ because there remain $3q$ free parameters after the targeting of A. Moreover, the regressors in the last column of $\Psi(\underline{\varepsilon_{t-1}})$ are zero, except when $i = j$ (lexicographical order).

3.3 The heterogenous portfolio model

Now, the underlying portfolios is composed of two homogeneous sub-portfolios whose dynamics behave differently. This situation is commonly met in finance, when several asset classes have to be managed simultaneously. The first (resp. second) portfolio corresponds to the assets that are numbered $\{1, \dots, p\}$ (resp. $\{p+1, \dots, N\}$). This necessitates to extend the previous model and to introduce more parameters. We need additional notations:

- For any real numbers $\alpha_1, \alpha_2, \alpha_3$, α_1 and α_3 being nonnegative, and two integers n and m, $n < m$, set the $m \times m$ matrix

  $$M(\alpha_1, \alpha_2, \alpha_3, m, n) := \begin{bmatrix} \alpha_1 e_n e_n' & \alpha_2 e_n e_{m-n}' \\ \alpha_2 e_{m-n} e_n' & \alpha_3 e_{m-n} e_{m-n}' \end{bmatrix}.$$

  By some standard algebraic calculations, we can prove that the characteristic polynomial of the symmetric matrix $M(\alpha_1, \alpha_2, \alpha_3, m, n)$ is

  $$x \mapsto (-1)^m x^{m-2} \left[(x - n\alpha_1)(x - (m-n)\alpha_3) - n(m-n)\alpha_2^2\right].$$

  Therefore, the associated spectrum is $\{x_+, x_-, 0\}$, $x_\pm := (n\alpha_1 + (m-n)\alpha_3 \pm$

$\sqrt{\Delta})/2$, where $\Delta := (n\alpha_1 + (m-n)\alpha_3)^2 - 4n(m-n)(\alpha_1\alpha_3 - \alpha_2^2) \geq 0$. These eigenvalues x_+ and x_- are nonnegative iff $\alpha_1\alpha_3 \geq \alpha_2^2$, and then the matrix $M(\alpha_1, \alpha_2, \alpha_3, m, n)$ is nonnegative. Note that this can be achieved in an optimization program with linear constraints by imposing that $\alpha_2 \leq \min(\alpha_1, \alpha_3)$.

- Set the partitioned matrix $\tilde{M}(\beta_1, \beta_2, \beta_3, p) = [\tilde{M}_{i,j}]_{1\leq i,j\leq N}$, where

$$\begin{aligned}\tilde{M}_{i,j} &= [\mathbf{1}((r,s)=(i,j)).\{\beta_1 \mathbf{1}(r \leq p, s \leq p) + \beta_3 \mathbf{1}(r > p, s > p) \\ &+ \quad \beta_2 \mathbf{1}(r \leq p, s > p) + \beta_2 \mathbf{1}(r > p, s \leq p)\}]_{1\leq r,s\leq N}\,.\end{aligned}$$

  By a similar reasoning as previously, it can be proved that the matrix $\tilde{M}(\beta_1, \beta_2, \beta_3, p)$ is nonnegative iff $\beta_1\beta_3 \geq \beta_2^2$. Again, it is sufficient that $\beta_2 \leq \min(\beta_1, \beta_3)$.

- Let γ_1 and γ_2 be two arbitrary nonnegative real numbers, and an integer $p \leq N$. Let $J := \{1, N+2, 2N+3, \ldots, (p-1)N+p\}$ and $\tilde{J} := \{pN+p+1, (p+1)N+p+2, \ldots, (N-1)N+N\}$. Set the diagonal matrix

$$\begin{aligned}N(\gamma_1, \gamma_2, p) &:= D\left(\gamma_1 e_{N^2,J} + \gamma_2 e_{N^2,\tilde{J}}\right) \\ &= \left[\mathbf{1}((r,s)=(i,j)).\left\{\gamma_1 \mathbf{1}(i=j\in J) + \gamma_2 \mathbf{1}(i=j\in \tilde{J})\right\}\right].\end{aligned}$$

  Obviously, $N(\gamma_1, \gamma_2, p)$ is nonnegative when γ_1 and γ_2 are nonnegative.

Now, let us define the "heterogeneous portfolio" model. With the notations above, we will choose the matrices B_k of (2.2) in the following parametric family:

$$\begin{aligned}\tilde{\mathcal{B}} = \{B \in \mathcal{M}_{N^2\times N^2}(\mathbb{R})\,|\, B = M(\alpha_1, \alpha_2, \alpha_3, N^2, Np) + \tilde{M}(\beta_1, \beta_2, \beta_3, p) + N(\gamma_1, \gamma_2, p),\\ \alpha_1 \geq 0, \alpha_3 \geq 0, \alpha_1\alpha_3 \geq \alpha_2^2, \beta_1 \geq 0, \beta_3 \geq 0, \beta_1\beta_3 \geq \beta_2^2, \gamma_1 \geq 0, \gamma_2 \geq 0\}.\end{aligned} \quad (3.2)$$

The non negativeness of such a $B \in \tilde{\mathcal{B}}$ is guaranteed when it is the case for the corresponding $M(\alpha_1, \alpha_2, \alpha_3, N^2, Np)$, $\tilde{M}(\beta_1, \beta_2, \beta_3, p)$ and $N(\gamma_1, \gamma_2, p)$.

To be more explicit, the latter model is defined by

$$\varepsilon_{it}\varepsilon_{jt} = a_{ij}+\sum_{k=1}^{q}\left((\alpha_{ij}^{(k)}+\beta_{ij}^{(k)}+\gamma_i^{(k)}\mathbf{1}(i=j))\varepsilon_{i,t-k}\varepsilon_{j,t-k}+\alpha_{ij}^{(k)}\sum_{(r,s)\neq(i,j)}\varepsilon_{r,t-k}\varepsilon_{s,t-k}\right)+\zeta_{ij,t},$$

$$\alpha_{i,j}^{(k)} = \alpha_1^{(k)}\mathbf{1}((i,j)\in J^2)+\alpha_3^{(k)}\mathbf{1}((i,j)\in\tilde{J}^2)+\alpha_2^{(k)}\mathbf{1}((i,j)\in J\times\tilde{J} \text{ or } (i,j)\in\tilde{J}\times J),$$

$$\beta_{i,j}^{(k)} = \beta_1^{(k)}\mathbf{1}((i,j)\in J^2)+\beta_3^{(k)}\mathbf{1}((i,j)\in\tilde{J}^2)+\beta_2^{(k)}\mathbf{1}((i,j)\in J\times\tilde{J} \text{ or } (i,j)\in\tilde{J}\times J),$$

$$\gamma_i^{(k)} = \gamma_1^{(k)}\mathbf{1}(i\in J)+\gamma_2^{(k)}\mathbf{1}(i\in\tilde{J}),$$

for any $k=1,\ldots,q$. This parametric model seeks to capture three effects on the dynamics of $\varepsilon_{i,t}\varepsilon_{j,t}$:

(i) a uniform effect of all past cross-products on every $\varepsilon_{i,t}\varepsilon_{j,t}$ through the coefficients $\alpha.$; when i and j belong to the first (resp. second) group of assets, we use α_1 (resp. α_3). When i and j do not belong to the same group, we invoke α_3.

(ii) a more important bump caused by the past values of $\varepsilon_{i,t}\varepsilon_{j,t}$ on itself, through the $\beta.$; as above, such effects depend on the group of i and j.

(iii) an additional bump when variances are managed (ie when $i=j$) through the parameters $\gamma.$; if i belongs to the first or the second group of assets, we apply γ_1 or γ_2 respectively.

Actually, the latter heterogeneous model specification can be criticized because the effect of $\varepsilon_{r,t-k}\varepsilon_{s,t-k}$ on $\varepsilon_{i,t}\varepsilon_{j,t}$, $(r,s)\neq(i,j)$, is transmitted through the same coefficient $\alpha_{ij}^{(k)}$, independently of the identify of the (r,s)-group. For instance, it is likely that this effect should be stronger when (r,s) and (i,j) belong to the same subset, typically. Therefore, a more general parametric model could be considered, where there would exist different cross-effects on the dynamics of $\varepsilon_{i,t}\varepsilon_{j,t}$, depending on the considered couples of indices (r,s), with our previous notations.

This so-called "extended heterogeneous model" would be the same as previously, except that the matrices $M(\cdot)$ have to be chosen differently. To be specific, instead of choosing $M(\alpha_1, \alpha_2, \alpha_3, N^2, Np)$ to build an element of $\tilde{\mathcal{B}}$, we select a $N^2 \times N^2$-block matrix inside $\bar{\mathcal{M}} := \{\bar{M} = [\bar{M}_{i,j}]_{1\leq i,j\leq N}\}$, where the $N \times N$ matrices $\bar{M}_{i,j}$ are defined as

$$\bar{M}_{i,j} = M(\alpha_1^{(1)}, \alpha_2^{(1)}, \alpha_3^{(1)}, N, p) \text{ if } i \text{ and } j \text{ belong to the first group,}$$

$$\bar{M}_{i,j} = M(\alpha_1^{(2)}, \alpha_2^{(2)}, \alpha_3^{(2)}, N, p) \text{ if } i \text{ and } j \text{ belong to the second group, and}$$

$$\bar{M}_{i,j} = M(\delta_1, \delta_2, \delta_3, N, p) \text{ if } i \text{ and } j \text{ do not belong to the same group.}$$

This would enrich the flexibility and the realism of the model. Unfortunately, the calculation of the spectrum of matrices $\bar{M} \in \bar{\mathcal{M}}$ is difficult. And only highly non-linear conditions will be able to guarantee that such matrices will be nonnegative.

Nonetheless, we are convinced that it is valuable to study the impact of cross-effects on any product dynamics $\varepsilon_{i,t}\varepsilon_{j,t}$ differently. To stay tractable and with the same notations as above, we simplify the latter extended model by assuming that $\delta_1 = \delta_2 = \delta_3 := \delta$. This means that the effect of all past cross products of returns on the dynamics of $\varepsilon_{i,t}\varepsilon_{j,t}$ is uniform, when i and j do not belong to the same portfolio. Therefore, under this simplifying assumption, any matrix $\bar{M}$ in $\bar{\mathcal{M}}$ is written as

$$\bar{M}(\alpha^{(1)}, \alpha^{(2)}, \delta) := \begin{bmatrix} M(\alpha^{(1)}) & \cdots & M(\alpha^{(1)}) & M(\delta) & \cdots & M(\delta) \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ M(\alpha^{(1)}) & \cdots & M(\alpha^{(1)}) & M(\delta) & \cdots & M(\delta) \\ M(\delta) & \cdots & M(\delta) & M(\alpha^{(2)}) & \cdots & M(\alpha^{(2)}) \\ \vdots & \cdots & \vdots & \vdots & \cdots & \vdots \\ M(\delta) & \cdots & M(\delta) & M(\alpha^{(2)}) & \cdots & M(\alpha^{(2)}) \end{bmatrix}, \tag{3.3}$$

where

$$M(\alpha^{(1)}) := M(\alpha_1^{(1)}, \alpha_2^{(1)}, \alpha_3^{(1)}, N, p) \text{ appears } p^2 \text{ times in the upper left square,}$$

$M(\alpha^{(2)}) := M(\alpha_1^{(2)}, \alpha_2^{(2)}, \alpha_3^{(2)}, N, p)$ appears $(N-p)^2$ times in the lower right square, and

$$M(\delta) := \delta e_N e_N', \ \delta \in \mathbb{R}^+, \text{ appears } 2p(N-p) \text{ times.}$$

Proposition 3.1. *A matrix $\bar{M}$ defined as in (3.3) is nonnegative iff*

$$(\alpha_1^{(1)}, \alpha_3^{(1)}, \alpha_1^{(2)}, \alpha_3^{(2)}, \alpha_2^{(1)}, \alpha_2^{(2)}, \delta) \in \mathbb{R}_+^4 \times \mathbb{R}^3,$$

$$\Delta^{(k)} := \alpha_1^{(k)}\alpha_3^{(k)} - (\alpha_2^{(k)})^2 \geq 0, \ k = 1, 2, \textit{ and}$$

$$\Delta^{(1)}\Delta^{(2)} \geq \delta^2 \left(\alpha_1^{(1)} + \alpha_3^{(1)} - 2\alpha_2^{(1)}\right) \times \left(\alpha_1^{(2)} + \alpha_3^{(2)} - 2\alpha_2^{(2)}\right). \tag{3.4}$$

The latter condition (3.4) is nonlinear. Nonetheless, it is satisfied if $\alpha_2^{(k)} \leq \min(\alpha_1^{(k)}, \alpha_3^{(k)})$, $k = 1, 2$ and $\delta \leq \min(\alpha_2^{(1)}, \alpha_2^{(2)})/2$. Note that all the latter constraints are linear and can easily been taken into account in a convex optimization program.

Proof of Proposition 3.1. First let us study the positiveness of the quadratic form q_0 that is associated to the $pN \times pN$ symmetric matrix

$$B_0 = \begin{bmatrix} M(\alpha) & \cdots & M(\alpha) \\ \vdots & \cdots & \vdots \\ M(\alpha) & \cdots & M(\alpha) \end{bmatrix}, \tag{3.5}$$

where $\alpha = (\alpha_1, \alpha_2, \alpha_3)$. Let the two sets of indices

$\mathcal{I} := \{1, \ldots, p, N{+}1, \ldots, N{+}p, 2N{+}1, \ldots, 2N{+}p, \ldots, (p{-}1)N{+}1, \ldots, (p{-}1)N{+}p\}$, and

$\mathcal{J} := \{p{+}1, \ldots, N, N{+}p{+}1, \ldots, 2N, 2N{+}p{+}1, \ldots, 3N, \ldots, (p{-}1)N{+}p{+}1, \ldots, pN\}$.

Obviously, $\{1,\dots,pN\} = \mathcal{I}\cup\mathcal{J}$. Then, for any $\boldsymbol{x}\in\mathbb{R}^{pN}$,

$$\begin{aligned} q_0(\boldsymbol{x}) &= \alpha_1 \sum_{(i,j)\in\mathcal{I}^2} x_i x_j + \alpha_3 \sum_{(i,j)\in\mathcal{J}^2} x_i x_j + 2\alpha_2 \left(\sum_{i\in\mathcal{I}} x_i\right)\cdot\left(\sum_{j\in\mathcal{J}} x_j\right) \\ &= \alpha_1 \left(\sum_{i\in\mathcal{I}} x_i + \frac{\alpha_2}{\alpha_1}\sum_{j\in\mathcal{J}} x_j\right)^2 + \frac{\alpha_1\alpha_3-\alpha_2^2}{\alpha_1}\left(\sum_{j\in\mathcal{J}} x_j\right)^2. \end{aligned}$$

Therefore, the non-negativeness of q_0 (or B_0) is equivalent to $\alpha_1\ge 0$, $\alpha_3\ge 0$ and $\alpha_1\alpha_3\ge\alpha_2^2$.

Now, we consider the quadratic form q that is associated to $\bar{M}\in\bar{\mathcal{M}}$. Introduce

$$\mathcal{I}^* := \{1,\dots,N-p,N+1,\dots,2N-p,2N+1,\dots,3N-p,\dots,(N-p-1)N+1,\dots,(N-p-1)N+N-p\},$$

$$\mathcal{J}^* := \{N-p+1,\dots,N,2N-p+1,\dots,2N,3N-p+1,\dots,3N,\dots,(N-p-1)2N-p+1,\dots,(N-p)N\},$$

$$\tilde{\mathcal{I}} = \mathcal{I}^* + Np, \text{ and } \tilde{\mathcal{J}} = \mathcal{J}^* + Np,$$

with obvious notations. Note that $\{1,\dots,(N-p)N\} = \mathcal{I}^*\cup\mathcal{J}^*$, $\{Np+1,\dots,N^2\} = \tilde{\mathcal{I}}\cup\tilde{\mathcal{J}}$, and $\{1,\dots,N^2\} = \mathcal{I}\cup\mathcal{J}\cup\tilde{\mathcal{I}}\cup\tilde{\mathcal{J}}$. Set $y_1 := \sum_{i\in\mathcal{I}} x_i$, $y_2 = \sum_{i\in\mathcal{J}} x_i$, $y_3 := \sum_{i\in\tilde{\mathcal{I}}} x_i$ and $y_4 = \sum_{i\in\tilde{\mathcal{J}}} x_i$. By simple calculations, we get

$$\begin{aligned} q(\boldsymbol{x}) &= \alpha_1^{(1)} y_1^2 + \alpha_3^{(1)} y_2^2 + 2\alpha_2^{(1)} y_1 y_2 + \alpha_1^{(2)} y_3^2 + \alpha_3^{(2)} y_4^2 + 2\alpha_2^{(2)} y_3 y_4 + 2\delta(y_1+y_2)(y_3+y_4) \\ &= \alpha_1^{(1)}\left(y_1 + \frac{\alpha_2^{(1)}}{\alpha_1^{(1)}} y_2 + \frac{\delta}{\alpha_1^{(1)}}(y_3+y_4)\right)^2 + \frac{\Delta^{(1)}}{\alpha_1^{(1)}}\left(y_2 - \frac{\alpha_2^{(1)}\delta}{\Delta^{(1)}}(y_3+y_4)\right)^2 \\ &+ y_3^2\left(\alpha_1^{(2)} - \frac{\delta^2}{\alpha_1^{(1)}} - \frac{(\alpha_1^{(1)}-\alpha_2^{(1)})^2\delta^2}{\alpha_1^{(1)}\Delta^{(1)}}\right) + y_4^2\left(\alpha_3^{(2)} - \frac{\delta^2}{\alpha_1^{(1)}} - \frac{(\alpha_1^{(1)}-\alpha_2^{(1)})^2\delta^2}{\alpha_1^{(1)}\Delta^{(1)}}\right) \\ &+ 2y_3y_4\left(\alpha_2^{(2)} - \frac{\delta^2}{\alpha_1^{(1)}} - \frac{(\alpha_1^{(1)}-\alpha_2^{(1)})^2\delta^2}{\alpha_1^{(1)}\Delta^{(1)}}\right). \end{aligned}$$

Its non-negativeness is guaranteed when

$$\begin{aligned} &\left(\alpha_1^{(2)} - \frac{\delta^2}{\alpha_1^{(1)}} - \frac{(\alpha_1^{(1)}-\alpha_2^{(1)})^2\delta^2}{\alpha_1^{(1)}\Delta^{(1)}}\right)\times\left(\alpha_3^{(2)} - \frac{\delta^2}{\alpha_1^{(1)}} - \frac{(\alpha_1^{(1)}-\alpha_2^{(1)})^2\delta^2}{\alpha_1^{(1)}\Delta^{(1)}}\right) \\ &\quad\ge \left(\alpha_2^{(2)} - \frac{\delta^2}{\alpha_1^{(1)}} - \frac{(\alpha_1^{(1)}-\alpha_2^{(1)})^2\delta^2}{\alpha_1^{(1)}\Delta^{(1)}}\right)^2, \end{aligned}$$

providing the result after some simplifications. $\square$

Therefore, we propose a second family of parametric matrices B_k in the case of heterogenous portfolios (with two groups):

$$\bar{\mathcal{B}} = \{B \in \mathcal{M}_{N^2\times N^2}(\mathbb{R})\,|B = \bar{M}(\alpha^{(1)},\alpha^{(2)},\delta) + \tilde{M}(\beta_1,\beta_2,\beta_3,p) + N(\gamma_1,\gamma_2,p),$$
$$\alpha^{(j)} \in \mathbb{R}^3_+,\ j = 1,2, (\alpha^{(1)},\alpha^{(2)},\delta) \text{ satisfies the conditions of Proposition 3.1,}$$
$$\beta_1 \geq 0, \beta_3 \geq 0, \beta_1\beta_3 \geq \beta_2^2, \gamma_1 \geq 0, \gamma_2 \geq 0\}.$$

Therefore, we automatically obtain non-negative covariance matrices in such an "extended heterogeneous" (simplified) model.

The latter ideas can be extended by considering more than two homogeneous sub-portfolios, at the price of more notational and algebraic complexities.

3.4 Conditions of stationarity

The model dynamics are specified by the N^2 equations (2.4). Strictly speaking, they define a Vectorial Autoregressive model of order p and dimension N^2 (or $N(N+1)/2$ to avoid redundant equations). The vector of noises $(\vec{\zeta}_t)$ is a difference martingale. In other words, setting the N^2 vector $\vec{\boldsymbol{v}}_t = [\varepsilon_{it}\varepsilon_{jt}]_{(i,j)\in N^2}$, its dynamics is

$$\vec{\boldsymbol{v}}_t = A + \sum_{k=1}^{q} C_k \boldsymbol{v}_{t-k} + \vec{\zeta}_t,\ E_{t-1}[\vec{\zeta}_t] = 0, \tag{3.6}$$

where $C_k := [b_{ijk,rs}]_{\{(i,j),(r,s)\in N^2\}}$, with the previous notations. Obviously, there is a one-to-one mapping between $(C_1,\ldots,C_q)$ and $(B_1,\ldots,B_q)$. For instance, in the case of an homogeneous portfolio, the parametrization that we proposed in Subsection (3.1) induces the matrices $C_k := [\alpha_k + \beta_k \mathbf{1}((i,j) = (r,s)) + \gamma_k \mathbf{1}(i = j = r = s)]_{(i,j),(r,s)}$, $k = 1,\ldots,q$.

It is well-known that the system given by (3.6) has a unique strongly stationary solution when all complex number λ s.t.

$$det(\lambda^q I_{N^2} - \lambda^{q-1} C_1 - \ldots - \lambda C_{q-1} - C_q) = 0$$

satisfies $|\lambda| < 1$. See Hamilton (1994), for instance. Those λ are the eigenvalues of the $qN^2 \times qN^2$ matrix

$$M_C := \begin{bmatrix} 0_{N^2} & I_{N^2} & 0_{N^2} & \ldots & \ldots & 0_{N^2} \\ \vdots & 0_{N^2} & I_{N^2} & \ddots & \ldots & \vdots \\ \vdots & & \ddots & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & \ddots & 0_{N^2} \\ \vdots & & & & 0_{N^2} & I_{N^2} \\ C_q & C_{q-1} & \ldots & \ldots & \ldots & C_1 \end{bmatrix}.$$

Unfortunately, the calculations of M_C's eigenvalues in some particular cases rapidly show that the stationarity conditions are written as complex nonlinear functional of the model parameters.

For instance, when $q = 1$ and in the case of an homogeneous portfolio, the stationarity condition is equivalent to the following: the modulus of the eigenvalues of C_1 are strictly smaller than one. In this case, simple algebraic calculations show that the characteristic polynomial of M_C is

$$\chi(x) = (\beta+\gamma-x)^{N-1}(\beta-x)^{N^2-N-1}\left(x^2 - (N^2\alpha + 2\beta + \gamma)x + (N^2\alpha + \beta + \gamma)\beta + \alpha\gamma\right).$$

Its roots are strictly smaller than one iff

$$\beta + \gamma < 1, \text{ and } (N^2\alpha + \beta + \gamma)(1 - \beta) < 1 - \beta + \alpha\gamma. \tag{3.7}$$

The latter condition is nonlinear. Note that it is fulfilled if $N^2\alpha + \beta + \gamma < 1$. Moreover, when $N \to \infty$, (3.7) can be satisfied only if $\alpha(N)$ tends to zero as $O(1/N^2)$.

When $p = 2$, similar calculations allow the calculation of the characteristic polynomial of M_C, but its roots cannot be easily calculated analytically due to a four-order factor. Such analytic problems are exacerbated with larger p.

Despite that lack of explicit eigenvalues of M_C, some sufficient conditions for stationarity can be obtained. For instance, following Higham and Tisseur (2003) (Equation (2.12)), any eigenvalue λ of M_C satisfies

$$|\lambda| \leq \max\left(\frac{\|C_p\|_1}{\|C_{p-1}\|_1}, 2\frac{\|C_{k+1}\|_1}{\|C_k\|_1}, k = 1, \ldots, p-2\right),$$

where $\|M\|_1$ denotes the usual ℓ^1-matrix norm of any matrix M. In the case of our "homogeneous portfolio" model, $\|C_k\|_1 = N^2\alpha_k + \beta_k + \gamma_k$, and the latter sufficient condition means $N^2\alpha_{k+1}+\beta_{k+1}+\gamma_{k+1} \leq \frac{1}{2}(N^2\alpha_k+\beta_k+\gamma_k)$, for any $k = 1, \ldots, p-1$. In other words, we get stationarity when the autoregressive coefficients of successive lags decrease to zero exponentially fast with the lag index k.

Therefore, in general, it is difficult to explicitly introduce conditions of stationarity during the inference stage. Indeed, such conditions are hardly ever written as linear constraints. When this is the case, this is most often obtained through strong restrictions on the set of admissible parameters. This is why we recommend to check the stationarity of such ARCH processes (by numerically calculating the spectrum of M_C, for instance) ex post, after having estimated these models through a penalized OLS criterion.

3.5 The Cholesky-GARCH approach

Although the constraint free model of Subsection 3.1 is flexible, the uncertainty induced by some projections on the cone of nonnegative matrices cannot be easily evaluated. As for the previous homogeneous and heterogenous ARCH models, their parameters are constrained to obtain nonnegative matrices. Now, we present alternative dynamics whose driving parameters are not constrained, since the generated variance covariance matrices will be nonnegative by construction.

As in Darolles et al. (2017), we propose to invoke the Cholesky decomposition of H_t, i.e. $H_t = L_t G_t L_t'$, where L_t is lower triangular with ones on the diagonal, and G_t is diagonal. Set $G_t = \text{diag}(g_{i,t})$ and $L_t = [\ell_{ij,t}]$, where $\ell_{ij,t} = 0$ when $j > i$. The idea of the Cholesky-GARCH approach is to define the (H_t)-process by specifying the dynamics of (G_t) and (L_t). Set the random vectors $\boldsymbol{v}_t$ s.t. $\varepsilon_t := L_t \boldsymbol{v}_t$. Then, given $\mathcal{F}_{t-1}$, the components of $\boldsymbol{v}_t$ are uncorrelated: $Cov_{t-1}(\boldsymbol{v}_t) = G_t$. Note that $v_{1t} = \varepsilon_{1t}$ is "observable".

First, we set the dynamics of the conditional volatility of ε_{1t}: $\mathbb{E}[\varepsilon_{1t}^2 | \mathcal{F}_{t-1}] = \mathbb{E}[v_{1t}^2 | \mathcal{F}_{t-1}] = g_{1t}$, and assume an ARCH-type model $g_{1,t} = a_{1,0} + \sum_{k=1}^m a_{11,k} f_{k,t}$, where every random factor $f_{k,t}$ is $\mathcal{F}_{t-1}$-measurable, for some nonnegative constants $a_{1,0}, a_{11,k}$, $k = 1, \ldots, m$. Typically, the factors f_{kt} are functions of $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$ and of some of their cross-products. For instance, we will assume that

$$g_{1,t} = a_{1,0} + \sum_{k=1}^m \sum_{j=1}^N a_{11,jk} \varepsilon_{j,t-k}^2, \tag{3.8}$$

for some nonnegative constants $a_{1,0}$ and $a_{11,jk}$. We can estimate the latter ARCH-type linear equation by penalized OLS, as the latter equation may be rewritten

$$\varepsilon_{1,t}^2 = a_{1,0} + \sum_{k=1}^m \sum_{j=1}^N a_{11,jk} \varepsilon_{j,t-k}^2 + \zeta_{11,t},\ \mathbb{E}[\zeta_{11,t} | \mathcal{F}_{t-1}] = 0.$$

Note that there are no auto-regressive lagged terms $g_{1,t-k}$, $k \geq 1$, on the r.h.s. of (3.8), so that we stay inside the ARCH family.

Moreover, for every $i > 1$, we have by definition

$$\varepsilon_{it} = \sum_{j=1}^{i-1} \ell_{ij,t} v_{jt} + v_{it}, \text{ or } v_{it} = -\sum_{j=1}^{i-1} \beta_{ij,t} \varepsilon_{jt} + \varepsilon_{it},$$

by introducing $L_t^{-1} := [-\beta_{ij,t}]$. Then, if $i > j$, we will assume

$$\beta_{ij,t} = a_{ij,0} + \sum_{k=1}^m a_{ij,k} f_{k,t},\ i > j. \tag{3.9}$$

We can estimate all the latter coefficients thanks to an ordinary least squares objective function. Indeed, we have

$$\varepsilon_{2t} = \beta_{21,t}\varepsilon_{1t} + v_{2t} = (a_{21,0} + \sum_{k=1}^{m} a_{21,k} f_{k,t})\varepsilon_{1t} + v_{2t}, \tag{3.10}$$

with v_{2t} is uncorrelated with $\varepsilon_{1t} = v_{1t}$, given $\mathcal{F}_{t-1}$. The latter property guarantees the consistency of OLS estimates of (3.10) and we get the dynamics of $(\beta_{12,t})$. A similar reasoning can be led for every couple (i,j), $i > j$, using the fact that v_{it} is uncorrelated with $\varepsilon_{1t}, \ldots, \varepsilon_{i-1,t}$, given $\mathcal{F}_{t-1}$. This provides the dynamics of the processes $(\beta_{ij,t})$ and then $(\ell_{ij,t})$, $i > j$. Note that we can now estimate any vector $\boldsymbol{v}_t$ by $\hat{L}_t^{-1}\varepsilon_t$. Contrary to Darolles et al. (2017), there are no lagged terms $\beta_{ij,t-k}$ of the r.h.s. of (3.9). While they propose QML-type procedures, possibly equation-by-equation but without penalization, we can rely on OLS or even penalized OLS, equation-by-equation.

Now, we evaluate the process (g_{2t}) by setting $\hat{v}_{2t} = \varepsilon_{2t} - \hat{\ell}_{12,t}\varepsilon_{1t}$, with obvious notations. Then, as above , we can assume a process as

$$g_{2,t} = a_{2,0} + \sum_{k=1}^{m} a_{22,k} f_{k,t}.$$

The corresponding linear regression is here

$$\hat{v}_{2t}^2 = a_{2,0} + \sum_{k=1}^{m} a_{22,k} f_{k,t} + \zeta_{22,t}, \ \mathbb{E}[\zeta_{22,t} | \mathcal{F}_{t-1}] \simeq 0.$$

The latter linear model can be estimated by penalized OLS, snd so on: iteratively, we estimate the processes (g_{it}), $i > 1$.

This latter procedure automatically generates non negative covariance matrices by construction. The necessary and sufficient conditions to get stationary solutions of (3.8) are provided by Darolles et al. (2017) for general Cholesky-GARCH specifications. Nonetheless, it seems impossible to explicitly take such conditions into account during the estimation stage.

To be able to compare the size of all these coefficients, it may be useful to normalize the vector of returns. For instance, by centering and normalizing any component of ε_t, using the unconditional volatility of every component and not by their conditional volatilities. Indeed, otherwise, this would induce some annoying constraints as $\sum_{j=1}^{i-1} \ell_{ij,t}^2 g_{j,t} + g_{i,t} = E_{t-1}[\varepsilon_{i,t}^2] = 1$, for every i.

4 Empirical study

In this section, we carry out a simulation study to explore the accuracy performance of sparse ARCH models. To do so, we consider three simulation settings, where we will compare some estimated variance-covariance processes to the true ones. Based on the DGP given by (2.4), and given some initial values, we simulate the successive values of MGARCH processes of size $N = 4$ for the experiments 1 and 2, and of size $N = 4, 6, 10$ for the experiment 3. We launch this procedure for $T = 10000$ and we consider 100 different variance covariance matrix patterns, as described below. Once a series is simulated, we estimate the model under different model assumptions: a scalar DCC, a homogeneous ARCH, a constraint free ARCH, a Cholesky ARCH and their penalized versions. The estimated parameters allow the calculation of successive variance covariance matrices, which are here $\hat{H}_t^{dcc}$ for the DCC model, $\hat{H}_t^{hom}$ (resp. $\hat{H}_t^{hom\star}$) for the homogeneous ARCH (resp. penalized homogeneous ARCH), $\hat{H}_t^{cf}$ (resp. $\hat{H}_t^{cf\star}$) for the constraint free ARCH (resp. penalized constraint free ARCH), and $\hat{H}_t^{cho}$ (resp. $\hat{H}_t^{cho\star}$) for the Cholesky ARCH (resp. penalized Cholesky ARCH).

The adaptive version of the Sparse Group Lasso estimator is implemented, where the first step estimator is the unpenalized OLS estimator. By a cross-validation (CV) procedure, we select the regularization parameter together with the system that determines the convergence rate of the regularization parameters to satisfy the oracle property. We emphasize that the standard CV developed for i.i.d. data can not be used in our time series framework. Such a procedure for penalized models is described in Hastie et al. (2015, Chapter 2), for instance. To fix this issue, we used the hv-CV procedure devised by Racine (2000), which consists in leaving a gap between the test sample and the training sample, on both sides of the test sample. The regularization parameters also should satisfy specific convergence rates to satisfy the oracle property, as detailed in Poignard (2016, Section 6).

The lags in the homogeneous, constraint free and Cholesky models are defined a priori as follows: in the experiments 1 and 2, $q = 10$ (resp. $q = 8$) for the homogeneous model (resp. for the constraint free and Cholesky models). As for the experiment 3, $q = 20$ (resp. $q = 10$) for the homogeneous model (resp. for the constraint free and Cholesky models). These choices of q are set a priori. We specified more lags for experience 3 since the simulation setting implies much more heterogeneity due to the significant number of parameters to recover. We did not significantly increase the number of lags for both the Cholesky and constraint free since they are already complex models.

We compare the true variance covariance processes and the estimated ones through the aforementioned models. To do so, we specify a matrix distance, namely the Frobenius norm, defined as $||A - B||_F := \sqrt{\text{Trace}((A-B)'(A-B))}$. We compute the previous norm for each t and for

$$A = R_t, \text{ and } B \in \{\hat{H}_t^{dcc}, \hat{H}_t^{hom}, \hat{H}_t^{hom\star}, \hat{H}_t^{cf}, \hat{H}_t^{cf\star}, \hat{H}_t^{cho}, \hat{H}_t^{cho\star}\}.$$

We take the average of those quantities over T = 10000 periods of time. We obtain an average gap for all those simulations as this procedure is repeated 100 times.

Simulated experiment 1. As a particular case of (2.4), we consider the following data generating process, a simple version of the homogeneous portfolio model:

$$\varepsilon_{it}\varepsilon_{jt} = a_{ij} + \sum_{k=1}^{q} \left((\alpha_k + \beta_k + \gamma_k \mathbf{1}(i=j))\varepsilon_{i,t-k}\varepsilon_{j,t-k} + \alpha_k \sum_{(r,s)\neq(i,j)} \varepsilon_{r,t-k}\varepsilon_{s,t-k} \right) + \zeta_{ij,t},$$

for any couple (i,j). All coefficients $(\alpha_k, \beta_k, \gamma_k)$ are set to zero except $(\alpha_4, \beta_4, \gamma_4)$.

Denoting $\omega = (\alpha_4, \beta_4, \gamma_4)$, we consider the grids

$$\omega^{(1)} = (0.001, 0.1, 0.2), \quad \omega^{(2)} = (0.005, 0.3, 0.1),$$
$$\omega^{(3)} = (0.01, 0.5, 0.1), \quad \omega^{(4)} = (0.01, 0.3, 0.2).$$

Beside, the nonnegative symmetric matrix A is simulated as follows: $A_{ij} \sim \mathcal{U}([-0.02, 0.02])$, $i \neq j$ and $A_{ii} \sim \mathcal{U}([0.1, 0.2])$ and we project this matrix on the cone of nonnegative matrices. For each fixed value $\omega^{(j)}$, $j = 1, \ldots, 4$, the matrix A is simulated 100 times along these lines. We remind that $q = 10$ for the homogeneous model and $q = 8$ for both the constraint free and Cholesky processes. See the results in Table 1.

We can highlight some interesting remarks from this simulation study. First, the DCC specification is outperformed by the competing models, especially by the homogeneous model, which is not surprising. Moreover, there is a gain in precision when applying a regularization procedure for the constraint free model, which is significantly parameterized: the penalized version outperforms the unpenalized version. This support the need of constraining the parameters when considering a large number of parameters, even when $N = 4$.

Simulated experiment 2. Here, we still consider that the "true" DGP is given by an homogeneous portfolio model. We set all coefficients $(\alpha_k, \beta_k, \gamma_k)$ to zero except for $k = 4, 5, 6$, where we consider different fixed values. A is parameterized as in the simulated experiment 1. Denote $\omega = (\alpha_2, \beta_2, \gamma_2, \alpha_3, \beta_3, \gamma_3, \alpha_4, \beta_4, \gamma_4)$ and consider the different grids

$$\begin{aligned} \omega^{(1)} &= (0.02, 0.2, 0.02, 0.01, 0.1, 0.01, 0.001, 0.01, 0.01), \text{ and} \\ \omega^{(2)} &= (0.001, 0.3, 0.05, 0.0005, 0.2, 0.02, 0.00001, 0.1, 0.01). \end{aligned}$$

For every fixed $\omega^{(j)}$, $j = 1, 2$, A is simulated 100 times.

We remind that $q = 10$ for the homogeneous model and $q = 8$ for both the constraint free and Cholesky processes. The results are detailed in Table 2.

The same remarks hold here as in simulated experiment 1.

Simulated experiment 3. In this experiment setting, we simulate a more general model (2.4). In matrix notation, it is written

$$H_t = A + \sum_{k=1}^{q} (I_N \otimes \varepsilon'_{t-k}) B_k (I_N \otimes \varepsilon_{t-k}).$$

Concerning this "true" DGP, we select $q = 5$ and $N = 4$. The $N^2 \times N^2$ matrices B_k are selected as $B_{ij,k} \sim \mathcal{U}([-0.2, 0.2])$ and $B_{ii,k} \sim \mathcal{U}([0.1, 0.15])$ such that they satisfy the positive definite, stationarity conditions. Moreover, we impose an "ordering" constraint: $|B_{ij,k}| \leq |B_{ij,k-1}|$ for $k = 2, \ldots, 5$ and any couple (i, j). As for the symmetric and positive definite matrix A, we define $A_{ij} \sim \mathcal{U}([-0.02, 0.02])$, $i \neq j$ and $A_{ii} \sim \mathcal{U}([0.1, 0.2])$. We consider two settings: in setting 1, he B_k, $k = 1, \ldots, 5$, and A matrices are independently drawn 100 times. Setting 2 is exactly similar, except we impose that B_1 and B_2 are null matrices now.

Remind that $q = 20$ for the homogeneous model and $q = 10$ for both the constraint free and Cholesky processes. See the results in Table 3.

These results emphasize the good performances of the constraint free and the Cholesky processes when the observed patterns are heterogeneous. The gain in precision is significant once the adaptive SGL regularization is applied. Not surprisingly, the DCC and the homogeneous are outperformed in this simulated framework.

5 Conclusion

We have proposed general multivariate ARCH model specifications that are linear with respect to the underlying parameters. These models can be estimated thanks to Ordinary Least Squares procedures. Moreover, it is possible to consider a large number of lagged values, in particular to approximate multivariate GARCH patterns. This can be managed through a regularization procedure, the Sparse Group Lasso penalty, that fosters sparsity both at a group level and within a group. Moreover, this regularization procedure satisfies the oracle property and identifies the right underlying sparse model.

By simulation, there are no clear results showing the ability of any variance covariance model to outperform the other ones in all circumstances. Our proposed ARCH models most often outperformed the usual DCC model, but it is difficult to establish a hierarchy among the models we introduced. Nonetheless, the constraint-free approach seems to provide the best results when no clear model structure (homogeneous/heterogeneous) is present. In the latter case, the use of the penalized criterion with a large number of lags significantly improves the results. More empirical work is surely necessary to evaluate the sensitivity of our estimators w.r.t. misspecification, the number of lags, the regularizing parameters, etc.

Acknowledgements: This research has been financially supported by the labex Ecodec ("Economics and Decision Sciences").

References

[1] Asai, M., M. McAleer, and J. Yu. 2006. Multivariate Stochastic Volatility: A Review. *Econometric Reviews* 25: 145-175.

[2] Bauwens, L., S. Laurent, and J.V.K. Rombouts. 2006. Multivariate GARCH models: a survey. *Journal of Applied Econometrics* 21: 79-109.

[3] Billio, M., M. Caporin, and M. Gobbo. 2006. Flexible Dynamic Conditional Correlation multivariate GARCH models for asset allocation. *Applied Financial Economics Letters* 2(2): 123-130.

[4] Darolles, S., C. Francq, and S. Laurent. 2017. Asymptotic of Cholesky-GARCH models and time-varying Conditional Beta. Working Paper.

[5] Engle, R. 2002. Dynamic Conditional Correlation. *Journal of Business & Economic Statistics* 20: 339-350.

[6] Engle, R.F. and K.F. Kroner. 1995. Multivariate Simultaneous Generalized ARCH. *Econometric Theory* 11: 122-150.

[7] Fan, J., and R. Li. 2001. Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association* 96: 1348-1360.

[8] Fan, J., Y. Fan, and J. Lv. 2008. Large dimensional covariance matrix estimation using a factor model. *Journal of Econometrics* 147: 186197.

[9] Francq, C., and J.-M. Zakoïan. 2010. *GARCH models structure, statistical inference and financial applications.* John Wiley and Sons, Chichester, West Sussex, U.K.

[10] Gouriéroux, C. 1997. *ARCH Models and Financial Applications.* Springer.

[11] Hamilton, J.D. 1994. *Time Series Analysis.* Pinceton U.P.

[12] Hastie,T., R. Tibshirani, and M. Wainwright. 2015. *Statistical Learning with Sparsity: The Lasso and Generalizations.* Monographs on Statistics and Applied Probability 143. Chapman and Hall.

[13] N.J. Higham, and F. Tisseur. 2001. Bounds for Eigenvalues of Matrix Polynomials. *Linear Algebra and its Applications* 358: 5-22.

[14] Patton, A.J., and K. Sheppard. 2009. "Evaluating Volatility and Correlation Forecasts." In T. Mikosch, J.-P. Kreib, R.A. Davis, and T.G. Andersen (eds.), *Handbook of Financial Time Series.* Springer Berlin Heidelberg, 801-838.

[15] Poignard, B. 2016. "Asymptotic theory of the Sparse Group Lasso." Working paper Crest, arXiv:1611.06034.

[16] Racine, J. 2000. Consistent cross-validatory model-selection for dependent data: hv-block cross-validation. *Journal of Econometrics* 99: 39-61.

[17] Simon, N., J. Friedman, T. Hastie, and R. Tibshirani. 2013. A sparse group lasso. *Journal of Computational and Graphical Statistics.* 22(2): 231-245.

[18] Tibshirani, R., and X. Suo. 2016. An Ordered Lasso and Sparse Time-Lagged Regression. *Technometrics* 4: 415-423.

Table 1: Average distance between true and estimated covariance matrices. The true DGP is the so-called “simulated experiment 1”. The estimated models are indicated at the top of every column (the penalized versions are denoted with a star)

ω	$\hat{H}_t^{dcc}$	$\hat{H}_t^{hom}$	$\hat{H}_t^{hom\star}$	$\hat{H}_t^{cf}$	$\hat{H}_t^{cf\star}$	$\hat{H}_t^{cho}$	$\hat{H}_t^{cho\star}$
$\omega^{(1)}$	0.2015	0.0776	0.1540	0.1042	0.0816	0.1516	0.1657
$\omega^{(2)}$	0.4647	0.1497	0.3117	0.1514	0.1401	0.3219	0.3346
$\omega^{(3)}$	1.2292	0.5341	0.7675	0.5063	0.3983	0.8386	0.8486
$\omega^{(4)}$	0.7353	0.2782	0.3545	0.2378	0.2047	0.4157	0.4350

Table 2: Average distance between true and estimated covariance matricesThe true DGP is the so-called "simulated experiment 2". The estimated models are indicated at the top of every column (the penalized versions are denoted with a star).

ω	$\hat{H}_t^{dcc}$	$\hat{H}_t^{hom}$	$\hat{H}_t^{hom\star}$	$\hat{H}_t^{cf}$	$\hat{H}_t^{cf\star}$	$\hat{H}_t^{cho}$	$\hat{H}_t^{cho\star}$
$\omega^{(1)}$	0.4914	0.2512	0.4095	0.2079	0.1537	0.4488	0.4503
$\omega^{(2)}$	0.9787	0.5209	0.7895	0.3669	0.3364	0.7658	0.7812

Table 3: Average distance between true and estimated covariance matrices. The true DGP is the so-called "simulated experiment 3". The estimated models are indicated at the top of every column (the penalized versions are denoted with a star).

	$\hat{H}_t^{dcc}$	$\hat{H}_t^{hom}$	$\hat{H}_t^{hom\star}$	$\hat{H}_t^{cf}$	$\hat{H}_t^{cf\star}$	$\hat{H}_t^{cho}$	$\hat{H}_t^{cho\star}$
Setting 1	0.4044	0.4181	0.4457	0.3780	0.2024	0.2833	0.2251
Setting 2	0.2870	0.2875	0.2952	0.1979	0.1121	0.1688	0.1440