High-dimensional penalized ARCH processes

Benjamin Poignard (Osaka University)[*],

Jean-David Fermanian (Ensae-Crest)[†]

May 2020

[*]E-mail address: bpoignard@econ.osaka-u.ac.jp (corresponding author)

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

Abstract

We introduce a general methodology to consistently estimate multidimensional ARCH models equation-by-equation, possibly with a very large number of parameters through penalization (sparse group-lasso). Some families of multidimensional ARCH models are proposed to tackle homogeneous or heterogeneous portfolios of assets. The corresponding conditions of stationarity and of positive definiteness are studied. We evaluate the relevance of such a strategy by simulation. The relative forecasting performances of our models are compared through the management of financial portfolios.

JEL classification: C13, C32, G17.

Keywords: Multivariate ARCH, Positive definiteness, Sparse group lasso, Stationarity.

1 Introduction

Modelling 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 purposes.

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 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 modelling correlation processes, the Quadratic Flexible DCC of Billio and Caporin (2006), among others. However, it would be unrealistic to capture heterogeneous 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, which is a strong assumption.

Another approach is given by factor modelling, 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 approach requires the identification of the 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 modelling high-dimensional variance-covariance matrices within the multivariate GARCH framework in a flexible manner 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 pa-

rameters, 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. (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 2018). 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 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 Cholesky-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 properties (positiveness, stationarity). In Section 4, we compare the performances of our penalized multivariate ARCH processes with other competitors by simulation

and real data.

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. We decompose r_t as the sum of its conditional expected return and a residual:

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

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 by the returns r_{t-k}, $k = 0, 1, 2, \ldots$. We set $H_t(\theta) = \mathrm{Var}(r_t|\mathcal{F}_{t-1}) := \mathrm{Var}_{t-1}(r_t) = \mathrm{Var}_{t-1}(\varepsilon_t)$ the $N \times N$ square symmetric and positive definite variance covariance 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 moment. 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 centered $\mathbb{E}[\eta_t] = 0$ with unit variance. 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 the row vectors $C_{ij,l}$, $i, j \in \{1, \ldots, N\}$, $k = 1, \ldots, q$ 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 definite 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 $B_k := [B_{ijk}]_{1 \leq i,j \leq N}$ 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 definite.

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 definite matrices B, due to their sizes ($qN^2(N^2+1)/2$ unknown parameters!) and due to the tedious nonlinear constraints imposed by non-negative definiteness (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

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

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

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

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. $i \leq j$.

Therefore, we will focus on the family of ARCH-type models given by the equations (2.4). For the sake of inference, we will assume such multivariate process (ε_t) is stationary. But it is difficult to exhibit necessary and/or sufficient conditions for stationarity that could be explicitly written. See a discussion in Section 6.1 in the appendix.

The previous linear model will be estimated by a penalized least squares pro-

cedure. 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 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

$$\left\{\begin{array}{lcl} \mathbb{G}_T l(\theta) & = & \dfrac{1}{T}\sum_{t=1}^{T} l(\varepsilon_t;\theta), \\ l(\varepsilon_t;\theta) & = & \|\mathrm{Vech}(\varepsilon_t\varepsilon_t') - \Psi(\underline{\varepsilon_{t-1}})\theta\|_2^2, \end{array}\right. \tag{2.5}$$

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

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

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

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

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

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 ob-

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

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

with $\alpha_{T,i}^{(k)} = |\tilde{\theta}_{T,i}^{(k)}|^{-\eta}$ and $\xi_{T,l} = \|\tilde{\theta}_T^{(l)}\|_2^{-\mu}$, where $\eta > 0, \mu > 0$, and $\tilde{\theta}_T$ 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, 2018).

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.

The asymptotic properties of the estimator $\hat{\theta}$ given in (2.6) are particular cases of the results in Poignard (2018). For the sake of completeness, we formally state the consistency and the asymptotic behavior of $\hat{\theta}$ in Section 6.1 in the appendix.

Obtaining the non-negativene definiteness 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 essentially from an empirical point of views since it hampers fast solving and closed-form algorithms. Thus, in Section 3, we propose parameterisations that allow for generating non-negative definite 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-negative definiteness 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 positive definite 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[1].

3 Our ARCH-type specifications

In this section, we propose several ARCH-type parameterisations of (2.2) to ensure the non negativeness of H_t. Remember 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 "heterogeneous" parametrization that is adapted to random vectors with discordant patterns. Finally, a model based on Cholesky 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 definite, onto $\mathcal{M}^+_{N\times N}(\mathbb{R})$, the cone of non-negative definite matrices. This method allows flexibility because one can independently specify and estimate the processes that are associated to each component of $\mathrm{vec}(\varepsilon_t\varepsilon_t')$. We rewrite the general dynamics given by (2.4) for each component of the $\varepsilon_t\varepsilon_t'$ matrix as

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

[1] Alternatively, we can invoke a parametrization of A in the cone of non-negative matrices. The natural basis would be provided by the spectral decomposition of $\mathbb{E}[\varepsilon_t\varepsilon_t']$ (or its empirical approximation $[\widehat{\mathrm{cov}}_{i,j}]$ instead). Indeed $\exists(\boldsymbol{v}_1,\ldots,\boldsymbol{v}_N)\in\mathbb{R}^N$ s.t. $\mathbb{E}[\varepsilon_t\varepsilon_t']\simeq[\widehat{\mathrm{cov}}_{i,j}]_{1\le i,j\le N}=\sum_{l=1}^N \nu_l\boldsymbol{v}_l\boldsymbol{v}_l'$, where $(\nu_1,\ldots,\nu_N)$ is the associated spectrum, $\nu_1\ge\nu_2\ge\cdots\ge\nu_N\ge 0$. Then, we could assume that there exist nonnegative real numbers π_l, $l=1,\ldots,N$ s.t. $A=\sum_{l=1}^N\pi_l\boldsymbol{v}_l\boldsymbol{v}_l'$. This allows to replacing the $N(N+1)/2$ unknown coefficients of A by N parameters $(\pi_1,\ldots,\pi_N)$.

if $i \leq j$. Through inference by OLS, the symmetric matrices A and B are not necessarily non-negative definite. 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 of 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 consider a portfolio of assets whose dynamics are more or less "similar": the influence of past returns on

current returns is comparable across all assets, because they similarly react to the same news/shocks; moreover, the influence of one asset on another one is roughly independent of the considered couple of assets. This is the so-called situation of an "homogeneous portfolio". For instance, this should be the situation for a portfolio of stocks chosen in the same country and/or industry.

Then, we will need some matrix notations:

- For any subset J of indices in $I := \{1, \dots, m\}$, 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 $\mathcal{J} = \{1, N+2, 2N+3, \dots, (N-2)N + N - 1, (N-1)N + N\}$, a subset of $\{1, \dots, N^2\}$. 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, $k = 1, \dots, q$, inside $\mathcal{B}$. More explicitly, in the homogeneous 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

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

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 heterogeneous portfolio model

Now, the underlying portfolio 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: bonds and stocks, bonds issued in the US or issued in Europe, etc. Therefore, the intra-group dynamics will share a certain degree of similarity, but they will be different from those between two assets that belong to different groups.

To be specific, the first (resp. second) portfolio corresponds to the assets that

are numbered $\{1,\ldots,p\}$ (resp. $\{p+1,\ldots,N\}$). 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:

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

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 $\alpha_\cdot$; 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_2.

(ii) a more important bump caused by the past values of $\varepsilon_{i,t}\varepsilon_{j,t}$ on itself, through the $\beta_\cdot$; 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 $\gamma_\cdot$; 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. The proof is reported in Appendix 6.1. □

Therefore, we propose a second family of parametric matrices B_k in the case of

heterogeneous 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 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 heterogeneous 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

$$g_{1,t} = a_{1,0} + \sum_{k=1}^{m} \sum_{j=1}^{N} a_{11,jk} \varepsilon_{j,t-k}^2, \tag{3.5}$$

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 (3.5), 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

$$\beta_{ij,t} = a_{ij,0} + \sum_{k=1}^{m} a_{ij,k} f_{k,t},\ i > j. \tag{3.6}$$

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

$$\varepsilon_{2t} = \beta_{21,t} \varepsilon_{1t} + v_{2t} = (a_{21,0} + \sum_{k=1}^{m} a_{21,k} f_{k,t}) \varepsilon_{1t} + v_{2t}, \tag{3.7}$$

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 (3.7) 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 (3.6). 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 (3.5) 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.

3.5 Extension to correlation processes

It is possible to apply the methodology we introduced in Section 2 to multivariate correlation processes. Let us study this idea, in the the case of DCC models. With the notations of Section 6.5 and as noticed by Aielli (2013), R_t is the conditional covariance of u_t and Q_t is the conditional covariance of $v_t := Q_t^{\star 1/2} u_t$. Then,

he proposed the so-called c-DCC versions of the previous models, where the (Q_t) dynamics are now defined as

$$Q_t = \bar{\Omega} + \sum_{k=1}^{p} \bar{M}_k Q_{t-k} \bar{M}_k' + \sum_{l=1}^{q} \bar{W}_l v_{t-l} v_{t-l}' \bar{W}_l', \text{ or} \tag{3.8}$$

$$Q_t = \bar{\Omega}^* + \sum_{k=1}^{p} \bar{B}_k \odot Q_{t-k} + \sum_{l=1}^{q} \bar{A}_l \odot v_{t-l} v_{t-l}'. \tag{3.9}$$

by denoting with a bar the new matrices of parameters. Since $Q_t = \mathbb{E}[v_t v_t' | \mathcal{F}_{t-1}]$, and assuming the auto-regressive components of the latter equations are zero, it is tempting to do the same reasoning as in Section 2. For instance, we can rewrite (3.9) as

$$v_t v_t' = \bar{\Omega}^* + \sum_{l=1}^{q} \bar{A}_l \odot v_{t-l} v_{t-l}' + \zeta_t,\ \mathbb{E}[\zeta_t | \mathcal{F}_{t-1}] = 0.$$

Then, the same penalized OLS strategy as above could be led by considering the random vectors (v_t) instead of the initial returns (ε_t). Unfortunately, v_t is not directly observable: $v_{i,t} = q_{ii,t}^{1/2} \varepsilon_{i,t} / \sqrt{h_{ii,t}}$ for any i, i.e. v_t depends on the conditional variances and on the diagonal terms of Q_t. The former quantities $h_{ii,t}$ can be estimated in a first step, typically by assuming univariate GARCH processes for every margin and yielding the standardized (estimated) returns $\hat{u}_t = \hat{D}_t^{-1} \varepsilon_t$. The latter quantities $q_{ii,t}$ are more annoying. They satisfy the (approximated) *nonlinear* dynamics

$$q_{ii,t} = \bar{\Omega}_{ii}^* + \sum_{l=1}^{q} \bar{A}_{ii,l} q_{ii,t-l} \hat{u}_{t-l} \hat{u}_{t-l}',\ i = 1, \ldots, N,$$

with obvious notations. During a second stage, the latter coefficients $\bar{\Omega}_{ii}^*$ and $A_{ii,l}$, $l = 1, \ldots, q$, can be estimated by N univariate QML optimizations, for example. This provides estimated values for $q_{ii,t}$ and then for v_t, denoted by $\hat{v}_t$. Then, as a third stage, we are able to evaluate the other coefficients of $\bar{\Omega}^*$ and $\bar{A}_l$, $l = 1, \ldots, q$ by penalized OLS, considering the linear equations

$$\hat{v}_{i,t} \hat{v}_{j,t} = \bar{\Omega}_{i,j}^* + \sum_{l=1}^{q} \bar{A}_{ij,l} \hat{v}_{i,t-l} \hat{v}_{j,t-l} + \hat{\zeta}_{ij,t},\ \mathbb{E}[\hat{\zeta}_{ij,t} | \mathcal{F}_{t-1}] = 0,$$

when $i < j$. Therefore, all the developments and model specifications of the previous subsections could be rewritten in this new correlation-based framework, replacing the cross-products $\varepsilon_{it}\varepsilon_{jt}$ by $\hat{v}_{i,t}\hat{v}_{j,t}$. Nonetheless, a precise investigation of the theoretical and empirical properties of penalized estimation methods is left for further research.

4 Empirical study

4.1 Simulation study

In this section, we empirically investigate the ability of the proposed penalization method to better capture complex variance covariance processes. We simulate the stochastic N-vectorial process (ε_t) based on two data generating processes: the multivariate ARCH and the BEKK processes. For the multivariate ARCH with q^* lags - M-ARCH(q^*) in the rest of the paper - case, we consider

$$\begin{cases} \varepsilon_t &= H_t^{1/2}\eta_t, \\ H_t &= \Omega + \sum_{k=1}^{q^*}(I_N \otimes \varepsilon'_{t-k})A_k(I_N \otimes \varepsilon_{t-k}), \end{cases}$$

where q^* is the number of lagged matrices being functions of ε_{t-k} and the $N^2 \times N^2$ square matrices A_k satisfy the stationarity conditions of Theorem 2 of Boussama (2006) together with the positivity condition given by Gouriéroux (1997). We generate the diagonal elements of A_k from a uniform distribution $\mathcal{U}([0.01, 0.05])$ and the off-diagonal ones from $\mathcal{U}([-0.01, 0.01])$ under the ordering constraint $\forall k \geq 2, \forall i, j, |A_{k,ij}| \leq |A_{k-1,ij}|$. As for the matrix Ω, the diagonal elements are simulated in $\mathcal{U}([0.1; 0.2])$ and the off-diagonals components in $\mathcal{U}([-0.01, 0.01])$. As for the BEKK process, the data generating process is based on

$$\begin{cases} \varepsilon_t &= H_t^{1/2}\eta_t, \\ H_t &= \Omega + A\varepsilon_{t-1}\varepsilon'_{t-1}A' + BH_{t-1}B', \end{cases}$$

where A, B are $N \times N$ matrices, satisfying the stationarity constraint $\|D_N^+\{(A \otimes A) + (B \otimes B)\}D_N\|_s < 1$, where D_N is the duplication matrix and D_N^+ the elemination matrix (see subsection 11.3 "Stationarity of VEC and BEKK Models" of Francq and Zakoïan (2010) for the stationarity condition and remark 11.1 for the definition of the latter matrices). The entries of A and B are generated among the uniform distribution $\mathcal{U}([-0.8, 0.8])$. The matrix Ω is generated as in the M-ARCH(q^*) case. Unlike the M-ARCH(q^*) case, the BEKK dynamic includes an autoregressive component through B, which motivated the use of larger lags when estimating our proposed parameterizations. In both proposed dynamics, we initialize the observations $(\varepsilon_k, \cdots, \varepsilon_1)$ with centered and unit variance multivariate Gaussian distribution, where $k = q^*$ in the M-ARCH model and $k = 1$ in the BEKK. Then conditionally on the past k observations, we generate H_t and thus ε_t according to a centered multivariate Gaussian distribution with variance covariance H_t.

We consider five problem sizes, $N = 10, 20, 30, 50, 100$, and $T = 5000$ observations for each of them. For the M-ARCH(q^*)-based data generating process, we considered different q^* depending on the problem size essentially due to the difficulty to satisfy the stationarity constraint. For $N = 10$, we selected $q^* = 2$; for $N = 20, 30, 50, 100$, we considered $q^* = 1$. Then we propose to compare the true variance covariance processes - BEKK and M-ARCH(q^*) - and the estimated ones through the constraint free, the Cholesky and the scalar DCC corresponding to process (6.5) with scalar matrix parameters. The estimation of the DCC model is based on the classic two-step Gaussian QMLE, where the marginal conditional volatility processes are specified as GARCH(1,1) and a correlation targeting procedure is applied in the second step, providing an estimated trajectory $\hat{H}_t^{dcc}$. Note that, as an alternative and following the procedure of Pakel, Shephard, Sheppard, and Engle (2018), we considered the composite likelihood based method for estimating the scalar DCC. To do so, we averaged across all pairs of standardized variables $(\hat{u}_{it}, \hat{u}_{jt})$ across the sample and computed the composite likelihood function corresponding to equation (3) of Pakel et al. (2018). We denote the resulting

covariance process by $\hat{H}_t^{cdcc}$.

Regarding our proposed variance-covariance dynamics, both constraint free and Cholesky models, denoted by $\hat{H}_t^{cf}$, $\hat{H}_t^{cho}$, together with their penalized counterparts, denoted by $\hat{H}_t^{cf\star}$, $\hat{H}_t^{cho\star}$ are considered when comparing the estimated variance covariance to the proposed true variance covariance. In the M-ARCH(q^*) case, when $N = 10$, we set $q = 5$ (resp. $q = 20$) for the constraint free (resp. Cholesky) models. For $N = 20, 30$, we set $q = 3$ (resp. $q = 5$) for the constraint free (resp. Cholesky) models. When $N = 50, 100$, we set $q = 2$ for both Cholesky and constrained free models. In the BEKK case, when $N = 10$, we set $q = 10$ (resp. $q = 30$) for the constraint free (resp. Cholesky) models. For $N = 20, 30$, we set $q = 8$ (resp. $q = 20$) for the constraint free (resp. Cholesky) models. Finally, when $N = 50, 100$, we consider $q = 5$ for the constrained free model and $q = 10$ for the Cholesky dynamic. For both simulated processes, the larger the dimension becomes, the more constrained the number of lags is. The significant parameterisation, especially in the constrained free case, fosters the use of a restricted number of lags. Moreover, in the estimation, we applied the first variance-covariance targeting procedure described in subsection 2.3.

We compare the true variance covariance processes and the estimated ones through the aforementioned models [2]. 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

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

We take the average of those quantities over $T = 5000$ periods of time. We obtain an average gap for all those simulations as this procedure is repeated 100 times.

The adaptive version of the Sparse Group Lasso (SGL) estimator is implemented, where the first step estimator is the unpenalized OLS estimator. The

[2] The homogeneous model has not been compared to the other specifications because it is not able to fairly compete with the others: if the true DGP is actually an homogeneous model, it highly outperforms the other ones; on the other side, this is the opposite under misspecification.

Groups are arbitrarily defined as the set of covariates belonging to a specific lags. For example, when $N = 30$, there are 5 groups of covariates when penalizing by SGL the Cholesky model. The penalization at a group level is performed through $\boldsymbol{p}_2(.)$ in (2.6), where the number of groups is denoted by m. In our simulations, since there are 5 lags for the Cholesky model, then $m = 5$ in $\boldsymbol{p}_2(.)$. By a cross-validation (CV) procedure - see e.g. Hastie and al. (2015, Chap. 2) - , we selected the regularization parameter and emphasize that the standard CV developed for i.i.d. data can not be used in our time series framework. 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 (2018, Section 6).

Clearly, the relevance of penalization reported in Table 1 for the M-ARCH(q^*) based DGP and Table 2 for the BEKK based DGP, Subsection 6.4, increases with the size N, for any method (constraint-free or Cholesky). Moreover, DCC models are always beaten with penalized criteria, whatever N, when it is never the case with unpenalized criteria. Besides, the Cholesky-GARCH is always significanly better than the constraint-free approach under the unpenalized framework, but their performances are comparable when adding penalizations. Moreover, the average distance becomes larger for all specifications in the BEKK case. Although the relative difference between our penalized specifications and the DCC is slightly smaller compared with the M-ARCH(q^*) case, essentially due to the presence of an autoregressive component, the increased number of lags for the constrained free and Cholesky GARCH still provides better performances.

4.2 Application to real data

To assess the relevance of the proposed penalized method, we propose a real data experiment. To do so, we compare the forecasting performances of the covariance matrices H_t for a portfolio of daily financial returns composed of the MSCI stock

index for the following 10 countries: Finland, France, Germany, Greece, Hong-Kong, Italy, Japan, The Netherlands, the United-Kingdom and the United-States. We focus on direct out-of-sample evaluation methods, which allow for pairwise comparisons. They test whether some of the variance covariance models provide better forecasts in terms of portfolio volatility behavior. Following the methodology of Engle and Colacito (2006), we develop a mean-variance portfolio approach to test the H_t forecasts. Intuitively, if a conditional covariance process is misspecified, then the minimum variance portfolio should emphasize such a shortcoming, compared to other models. Then, consider an investor who allocates a fixed amount between N stocks, according to a minimum-variance strategy and independently at each time t. At each date t, he/she solves

$$\min_{w_t} \; w_t' H_t w_t, \;\; \text{s.t.} \;\; \iota' w_t = 1, \tag{4.1}$$

where w_t is the $N \times 1$ vector of portfolio weights chosen at (the end of) time $t-1$, ι is a $N \times 1$ vector of 1 and H_t is the estimated conditional covariance matrix of the asset returns at time t. They are deduced from some dynamics that have been estimated on the sub-sample December 1998 - November 2015. Once the latter process is estimated in-sample, out-of-sample predictions are plugged into the program (4.1) between December 2015 and March 2018. The solution of (4.1) is given by the global minimum variance portfolio $w_t = H_t^{-1}\iota / \iota' H_t^{-1} \iota$.

Engle and Colacito (2006) show that the realized portfolio volatility is the smallest one when the variance covariance matrices are correctly specified. As a consequence, if wealth is allocated using two different dynamic models i and j, whose predicted covariance matrices are (H_t^i) and (H_t^j), the strategy providing the smallest portfolio variance will be considered as the best one. To do so, we consider a sequence of minimum variance portfolio weights $(w_{i,t})$ and $(w_{j,t})$, depending on the model. Then, we consider a distance based on the difference of the squared returns of the two portfolios, defined as $u_{ij,t} = \left\{ w_{i,t}' \epsilon_t \right\}^2 - \left\{ w_{j,t}' \epsilon_t \right\}^2$. The portfolio variances are the same if the predicted covariance matrices are the same. Thus

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

We run the latter test to compare the scalar DCC (DCC), the Orthogonal GARCH (O-G), the BEKK (BEKK), the constraint free (Cf), the Cholesky GARCH (Chol), the homogeneous (Hom), and their penalized counterpart (displayed with *). The definition of the BEKK and O-GARCH processes are reported in Subsection 6.5. For the homogeneous specification, we chose $q = 30$. As for the Cholesky dynamics, we selected $q = 15$ and $q = 30$ for the constraint free model. We apply the same variance-covariance targeting procedure as in the simulation setting. We applied the SGL regularization, where the groups were defined as the set of covariates belonging to a specific lag. The matrix forecast comparisons are provided in Table 3, Subsection 6.4.

These results first highlight the clear gain in prediction accuracy when penalizing a highly parameterized dynamic. The DCC, O-GARCH and BEKK are outperformed by the regularized Cholesky, constraint free and homogeneous processes. No clear hierarchy emerges between the three later methodologies. Note that, although the portfolio is composed with heterogeneous countries, the homogeneous variance covariance specification performs well.

This test for variance covariance specification provided some insights about which dynamic would be more relevant for the portfolio we aim at modelling, although no a priori criterion is at hand to choose the correct model. First, our results tend to emphasize that richly parameterised models are more fit to capture the patterns for heterogeneous variables. Second, our penalization procedure explicitly manages the over-fitting issue and aims at a balance between parsimony and yet sufficiently parameterised. These results are also in line with the simulation we carried out in the previous subsection: a high-dimensional variance covariance

process in the presence of heterogeneous patterns induced, e.g., by the multivariate BEKK, would confirm that a richly parameterised dynamic such as the constrained free - with a sensible number of lags - would be more suitable than a constrained version of the DCC.

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. Then, we have considered a large number of lagged values to approximate multivariate GARCH patterns, and they can be managed through a regularization procedure. To do so, the Sparse Group Lasso penalty is relevant as it fosters sparsity both at a group level and within a group. Besides, our multivariate ARCH framework is devised such that the penalized objective function is convex with convex constraints. The regularization procedure thus satisfies the oracle property and identifies the right underlying sparse model.

By simulation and with our empirical experiment, 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 are not outperformed by the DCC model in general. More interestingly, there is a gain in regularizing the estimates once the parameter vector size becomes significant, even for small vector sizes. Nonetheless, more empirical work is necessary to evaluate the sensitivity of our models 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"); and the Japan Society for the Promotion of Science.

References

[1] Aeilli, G.P. (2013). Dynamic Conditional Correlation: On Properties and Estimation. *Journal of Business & Economic Statistics* 31: 282-299.

[2] Alexander, C. (2001). Orthogonal GARCH in Alexander, C. (Ed.), *Mastering Risk*, Financial Times-Prentice Hall, London, pp. 21-28.

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

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

[5] 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.

[6] Boussama, F., Fuchs, F., Stelzer, R. (2011). Stationarity and geometric ergodicity of BEKK multivariate GARCH models. *Stochastic Processes and their applications* 121, 2331-2360.

[7] Darolles, S., C. Francq, and S. Laurent. (2018). Asymptotic of Cholesky-GARCH models and time-varying Conditional Beta. *Journal of Econometrics* 204: 223-247.

[8] Diebold, F.X., and R.S. Mariano. (1995). Comparing Predictive Accuracy. *Journal of Business & Economic Statistics* 13: 253-263.

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

[10] Engle, R., and R. Colacito. (2006). Testing and Valuing Dynamic Correlations for Asset Allocation. *Journal of Business & Economic Statistics* 24: 238-253.

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

[12] 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.

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

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

[15] Gouriéroux, C. (1997). *ARCH Models and financial Applications.* Springer.

[16] 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.

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

[18] Pakel, C., N. Shephard, K. Sheppard, and R. F. Engle (2018). Fitting vast dimensional time-varying covariance models. Mimeo.

[19] 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.

[20] Poignard, B. (2018). Asymptotic Theory of the Adaptive Sparse Group Lasso. *Annals of the Institute of Statistical Mathematics.* https://doi.org/10.1007/s10463-018-0692-7.

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

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

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

6 Appendix

6.1 Conditions of stationarity

The model dynamics are specified by the N^2 equations (2.4). They formally 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 martingale difference. In other words, setting the N^2 vector $\vec{\boldsymbol{v}}_t = [\varepsilon_{it}\varepsilon_{jt}]_{(i,j)\in N^2}$, the ARCH dynamics implies

$$\vec{\boldsymbol{v}}_t = A + \sum_{k=1}^{q} C_k \vec{\boldsymbol{v}}_{t-k} + \vec{\zeta}_t,\ \mathbb{E}_{t-1}[\vec{\zeta}_t] = 0, \tag{6.1}$$

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 $C_k := [\alpha_k + \beta_k \mathbf{1}((i,j) = (r,s)) + \gamma_k \mathbf{1}(i = j = r = s)]_{(i,j),(r,s)}$, $k = 1, \ldots, q$.

A usual necessary condition so that the system given by (6.1) has a strongly stationary is the following: any 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 Boussama et al. (2011), 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 latter 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

$$\beta + \gamma < 1, \text{ and } (N^2\alpha + \beta + \gamma)(1 - \beta) < 1 - \beta + \alpha\gamma. \tag{6.2}$$

The latter condition is nonlinear. Note that it is fulfilled if $N^2\alpha + \beta + \gamma < 1$. Moreover, when $N \to \infty$, (6.2) 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 stronger necessary 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, stationarity is "likely" when the autoregressive coefficients of successive lags decrease to zero exponentially fast with the lag index

k.

Therefore, in general, it is difficult to guarantee stationarity during the inference stage. Indeed, the usual necessary 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 whether such ARCH processes may be non-stationary (by numerically calculating the spectrum of M_C, for instance) ex post, after having estimated these models through a penalized OLS criterion.

6.2 Some asymptotic results

Here, we recall some asymptotic properties of the estimator $\hat{\theta}$. These results are stated in Poignard (2018, Section 4.2). Note that we do not cover the case of "covariance targeting", that is detailed in Section 2.3.

Let us denote by $\mathcal{A}$ the "true subset model", i.e. the set of indices that correspond to the nonzero coefficients of θ_0: $\mathcal{A} := \{j : \theta_{0,j} \neq 0\}$. Similarly for every group: $\mathcal{A}_k := \{(k,i) : \theta_{0,i}^{(k)} \neq 0\}$, $k \in \mathcal{S} := \{1, \dots, m\}$. The non-zero estimated parameters constitute the subset $\hat{\mathcal{A}} := \{i : \hat{\theta}_i \neq 0\}$. The usual oracle property will typically insure that the two subsets $\mathcal{A}$ and $\hat{\mathcal{A}}$ will coincide (in probability) for large T.

Theorem 6.1 (Consistency)**.** *Assume*

(i) (ε_t) *is a strictly stationary and ergodic process;*

(ii) the parameter set Θ *is convex (but not necessarily compact).*

Then, the sequence of penalized estimators $\hat{\theta}$ *satisfies*

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\| = O_p(T^{-1/2} + \lambda_T T^{-1} a_T + \gamma_T T^{-1} b_T),$$

where $a_T := card(\mathcal{A}).(\max_{k \in \mathcal{S}} (\max_{i \in \mathcal{A}_k} \alpha_{T,i}^{(k)}))$, $b_T := card(\mathcal{A}).(\max_{l \in \mathcal{S}} \xi_{T,l})$ *are some stochas-*

tic quantities such that $\lambda_T T^{-1} a_T \xrightarrow{\mathbb{P}} 0$ *and* $\gamma_T T^{-1} b_T \xrightarrow{\mathbb{P}} 0$.

The quantities $\alpha_{T,i}^{(k)}$ and $\xi_{T,l}$ are the adaptive weights entering the penalties defined in (2.6). The following theorem shows that the adaptive SGL satisfies the oracle property under proper convergence rates of λ_T and γ_T and provides the trade-off between the l^1/l^2 regularizer and the l^1 regularizer. Denote $\mathbb{H} := \mathbb{E}[\nabla^2_{\theta\theta'} l(\varepsilon_t; \theta_0)]$, $\mathbb{M} := \mathbb{E}[\nabla_\theta l(\varepsilon_t; \theta_0) \nabla_{\theta'} l(\varepsilon_t; \theta_0)]$ and $\mathbb{H}_{\mathcal{A}\mathcal{A}}, \mathbb{M}_{\mathcal{A}\mathcal{A}}$ are matrices taken over the set of active indices.

Theorem 6.2 (Asymptotic normality)**.** *Assume (i) and (ii) apply, and that*

(iii) $\mathbb{E}\Big[\big\{\|\varepsilon_t\|^2 + \|\Psi(\underline{\varepsilon_{t-1}})\|^2\big\}\|\Psi(\underline{\varepsilon_{t-1}})\|^2\Big] < \infty$;

(iv) $\lambda_T T^{-1/2} \to 0$, $\gamma_T T^{-1/2} \to 0$, $T^{(\eta-1)/2}\lambda_T \to \infty$, $T^{(\mu-1)/2}\gamma_T \to \infty$ *and* $T^{(\mu-\eta)/2}\gamma_T \lambda_T^{-1} \to \infty$ *where* $\mu, \eta > 0$;

(v) $\mathbb{H}$ *and* $\mathbb{M}$ *are definite positive.*

Then $\hat{\theta}$ *satisfies* $\lim_{T\to\infty} \mathbb{P}(\hat{\mathcal{A}} = \mathcal{A}) = 1$ *and*

$$\sqrt{T}(\hat{\boldsymbol{\theta}}_{\mathcal{A}} - \boldsymbol{\theta}_{0,\mathcal{A}}) \xrightarrow{d} \mathcal{N}(0, \mathbb{H}_{\mathcal{A}\mathcal{A}}^{-1} \mathbb{M}_{\mathcal{A}\mathcal{A}} \mathbb{H}_{\mathcal{A}\mathcal{A}}^{-1}).$$

6.3 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

$$B_0 = \begin{bmatrix} M(\alpha) & \cdots & M(\alpha) \\ \vdots & \cdots & \vdots \\ M(\alpha) & \cdots & M(\alpha) \end{bmatrix}, \tag{6.3}$$

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

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

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

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

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

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

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

with obvious notations. Note that $\{1, \ldots, (N{-}p)N\} = \mathcal{I}^* \cup \mathcal{J}^*$, $\{Np{+}1, \ldots, N^2\} = \tilde{\mathcal{I}} \cup \tilde{\mathcal{J}}$, and $\{1, \ldots, N^2\} = \mathcal{I} \cup \mathcal{J} \cup \tilde{\mathcal{I}} \cup \tilde{\mathcal{J}}$. 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

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

Its non-negativeness is guaranteed when

$$\left(\alpha_1^{(2)} - \frac{\delta^2}{\alpha_1^{(1)}} - \frac{(\alpha_1^{(1)} - \alpha_2^{(1)})^2\delta^2}{\alpha_1^{(1)}\Delta^{(1)}}\right) \times \left(\alpha_3^{(2)} - \frac{\delta^2}{\alpha_1^{(1)}} - \frac{(\alpha_1^{(1)} - \alpha_2^{(1)})^2\delta^2}{\alpha_1^{(1)}\Delta^{(1)}}\right)$$
$$\geq \left(\alpha_2^{(2)} - \frac{\delta^2}{\alpha_1^{(1)}} - \frac{(\alpha_1^{(1)} - \alpha_2^{(1)})^2\delta^2}{\alpha_1^{(1)}\Delta^{(1)}}\right)^2,$$

providing the result after some simplifications.

6.4 Tables

Table 1: Average distance true/estimated covariance matrices - M-ARCH(q^*) case

	$\hat{H}_t^{dcc}$	$\hat{H}_t^{cdcc}$	$\hat{H}_t^{cf\star}$	$\hat{H}_t^{cf}$	$\hat{H}_t^{cho\star}$	$\hat{H}_t^{cho}$
$N = 10$	2.05	1.98	1.64	5.45	1.69	6.72
$N = 20$	6.82	6.51	5.59	33.68	5.67	14.43
$N = 30$	17.97	17.36	13.18	118.89	11.14	29.54
$N = 50$	25.18	24.98	19.28	355.61	13.24	81.04
$N = 100$	131.16	130.87	62.62	847.23	48.57	789.72

Table 2: Average distance true/estimated covariance matrices - BEKK case

	$\hat{H}_t^{dcc}$	$\hat{H}_t^{cdcc}$	$\hat{H}_t^{cf\star}$	$\hat{H}_t^{cf}$	$\hat{H}_t^{cho\star}$	$\hat{H}_t^{cho}$
$N = 10$	9.52	9.17	6.40	11.06	7.58	14.45
$N = 20$	38.75	38.28	26.03	40.87	27.45	61.22
$N = 30$	51.54	50.94	45.90	161.11	47.61	139.54
$N = 50$	88.04	87.86	76.42	237.61	78.60	274.20
$N = 100$	155.13	153.91	121.77	255.84	127.31	472.97

6.5 Some competing M-GARCH models

The BEKK model directly generates a variance covariance process. Developed by Baba, Engle, Kraft and Kroner, in a preliminary version of Engle and Kroner

Table 3: Diebold Mariano Test of Multivariate GARCH models

	DCC	O-G	BEKK	Hom	Hom*	Chol	Chol*	Cf	Cf*
DCC		$\mathbf{-3.943}^{c}$	-0.745	-1.040	$\mathbf{1.753}^{b}$	$\mathbf{-3.208}^{c}$	$\mathbf{1.875}^{b}$	$\mathbf{-3.889}^{c}$	$\mathbf{2.476}^{c}$
O-G	$\mathbf{3.943}^{c}$		$\mathbf{3.749}^{c}$	-1.036	$\mathbf{4.469}^{c}$	$\mathbf{-3.431}^{c}$	$\mathbf{2.179}^{b}$	$\mathbf{-3.155}^{c}$	$\mathbf{4.492}^{c}$
BEKK	0.745	$\mathbf{-3.749}^{c}$		-1.040	$\mathbf{2.382}^{c}$	$\mathbf{-3.888}^{c}$	$\mathbf{1.948}^{b}$	$\mathbf{-3.851}^{c}$	$\mathbf{2.085}^{b}$
Hom	1.040	1.036	1.040		1.041	1.008	1.036	1.030	1.042
Hom*	$\mathbf{-1.753}^{b}$	$\mathbf{-4.469}^{c}$	$\mathbf{-2.382}^{c}$	-1.041		$\mathbf{-3.972}^{c}$	-1.207	$\mathbf{-4.117}^{c}$	-1.189
Chol	$\mathbf{3.208}^{c}$	$\mathbf{3.431}^{c}$	$\mathbf{3.888}^{c}$	-1.008	$\mathbf{3.972}^{c}$		$\mathbf{3.519}^{c}$	$\mathbf{1.694}^{b}$	$\mathbf{3.964}^{c}$
Chol*	$\mathbf{-1.875}^{b}$	$\mathbf{-2.179}^{b}$	$\mathbf{-1.948}^{b}$	-1.036	1.207	$\mathbf{-3.519}^{c}$		$\mathbf{-3.955}^{c}$	$\mathbf{1.450}^{a}$
Cf	$\mathbf{3.889}^{c}$	$\mathbf{3.155}^{c}$	$\mathbf{3.851}^{c}$	-1.030	$\mathbf{4.118}^{c}$	$\mathbf{-1.694}^{b}$	$\mathbf{3.955}^{c}$		$\mathbf{4.126}^{c}$
Cf*	$\mathbf{-2.476}^{c}$	$\mathbf{-4.492}^{c}$	$\mathbf{-2.085}^{b}$	-1.041	1.189	$\mathbf{-1.450}^{a}$	$\mathbf{-1.450}^{c}$	$\mathbf{-4.126}^{c}$	

This table reports the out-of-sample t-statistics of the Diebold-Mariano test that checks the equality between covariance matrix forecasts using the loss function $u_{ij,t}$ over the period December 2015 and March 2018. This loss function is defined as the difference between squared realized returns of alternative MGARCH models. When the null hypothesis of equal predictive accuracy is rejected, a positive number is evidence in favor of the model in the column. a, b, c: rejection of the null hypothesis at 10%, 5% and 1% respectively.

(1995), the BEKK is specified as

$$\begin{cases} \varepsilon_t &= H_t^{1/2}\eta_t, \text{ with } H_t := \mathbb{E}[\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}] \succ 0 \text{ so that} \\ H_t &= \Omega + \sum_{k=1}^{q}\sum_{j=1}^{K} A_{kj}\varepsilon_{t-k}\varepsilon_{t-k}'A_{kj}' + \sum_{i=1}^{r}\sum_{i=1}^{K} B_{ij}H_{t-i}B_{ij}', \end{cases}$$

where K is an integer, Ω, A_{kj} and B_{kj} are square $N \times N$ matrices and $\Omega \succ 0$. One advantage of the BEKK model is there is no positive semi-definite constraint on the A_{kj} and B_{kj} matrices. However, it imposes highly artificial constraints on the volatilities and covariances of the components. As a consequence, the coefficients of a BEKK representation are difficult to interpret. In our application, a scalar BEKK was considered, where A_{kj} and B_{kj} are scalar with $K = 1$, $q = r = 1$, together with a Gaussian QMLE estimation.

Beside BEKK type dynamics, factor models provide rather natural alternatives. The O-GARCH assumes the decomposition $H_t = P\Lambda_t P'$, where $\Lambda_t = \text{diag}(\lambda_{1,t}, \cdots, \lambda_{K,t})$, with K the number of factors. Here, we choose $K = N$ factors and each λ_t is supposed to follow a univariate GARCH(1,1) process that is esti-

mated by maximum likelihood. The matrix P is nonsingular and it is estimated by PCA on the empirical variance covariance matrix of ϵ_t: see Alexander (2001), e.g.

Beside the latter direct specification of the covariance matrices (H_t) dynamics, an alternative road is to split the task into two parts: individual volatility dynamics on one side, and correlation dynamics on the other side. The most commonly used correlation process is the Dynamic Conditional Correlation (DCC) of Engle (2002). In its BEKK form, the general DCC model is specified as

$$\left\{\begin{array}{rcl} \varepsilon_t &=& H_t^{1/2}\eta_t, \text{ with } H_t := \mathbb{E}[\varepsilon_t\varepsilon_t'|\mathcal{F}_{t-1}] \succ 0 \text{ so that} \\ H_t &=& D_tR_tD_t,\ R_t = Q_t^{\star-1/2}Q_tQ_t^{\star-1/2}, \\ Q_t &=& \Omega + \sum_{k=1}^{p} M_kQ_{t-k}M_k' + \sum_{l=1}^{q} W_lu_{t-l}u_{t-l}'W_l', \end{array}\right. \tag{6.4}$$

where $D_t = \text{diag}\left(\sqrt{h_{11,t}}, \sqrt{h_{22,t}}, \ldots, \sqrt{h_{NN,t}}\right)$, $u_t = (u_{1,t}, \ldots, u_{N,t})$ with $u_{i,t} = \varepsilon_{i,t}/\sqrt{h_{ii,t}}$, $Q_t = [q_{ij,t}]$, $Q_t^\star = \text{diag}\,(q_{11,t}, q_{22,t}, \ldots, q_{NN,t})$. The model is parameterized by some deterministic matrices $(M_k)_{k=1,\cdots,p}$, $(W_l)_{l=1,\cdots,q}$ and a positive definite $N \times N$ matrix Ω. Alternatively, Engle (2002) considered a VEC-type specification too. Denoting by $\odot$ the Hadamard matrix product, the (Q_t)-dynamics become

$$Q_t = \Omega^* + \sum_{k=1}^{p} B_k \odot Q_{t-k} + \sum_{l=1}^{q} A_l \odot u_{t-l}u_{t-l}', \tag{6.5}$$

where the deterministic matrices $(B_k)_{k=1,\cdots,p}$ and $(A_l)_{l=1,\cdots,q}$ must be positive semi-definite.

Since the number of parameters of the latter models is of order $O(N^2)$, the matrices M_k and W_l (resp. B_k's and A_l) are often assumed to be scalar. This is typically a strong and questionable constraint, particularly when the dimension N increases or when the variables in (ε_t) are heterogeneous. Furthermore, their inference is usually carried out trough the QML method, based on a Gaussian or Student quasi likelihood function. Under this methodology, applying a regularization method, even possible, is numerically arduous and no general asymptotic results exist in this case (to the best of our knowledge), due to the non-convexity

of the QML criterion.