PROXIMITY-STRUCTURED MULTIVARIATE VOLATILITY MODELS

MASSIMILIANO CAPORIN° AND PAOLO PARUOLO♮

Abstract. In many multivariate volatility models, the number of parameters increases faster than the cross-section dimension, hence creating a curse of dimensionality problem. This paper discusses specification and identification of structured parameterizations based on weight matrices induced by economic proximity. It is shown that structured specifications can mitigate or even solve the curse of dimensionality problem. Identification and estimation of structured specifications are analyzed, rank and order conditions for identification are given and the specification of weight matrices is discussed. Several structured specifications compare well with alternatives in modelling conditional covariances of six returns from the New York Stock Exchange.

Date: First version: November 3, 2008. Current version: December 7, 2012.

JEL classification: C31, C32, G11

Running head: Proximity-Structured Volatility Models.

Key words and phrases. GARCH, Stochastic Volatility, Realized Volatility, Spatial models, Weight matrices.

° Università degli Studi di Padova, (I).

♮ European Commission Joint Research Centre, Institute for the Protection and Security of the Citizen & Department of Economics, University of Insubria, Varese (I), (corresponding author).

Corresponding address: Paolo Paruolo,
Department of Economics, University of Insubria,
Via Monte Generoso 71, 21100 Varese, Italy,
Email: paolo.paruolo@uninsubria.it, homepage: `http://eco.uninsubria.it/webdocenti/pparuolo/` .

Contents

1. Introduction

Multivariate volatility models (MVM) are used for asset pricing, portfolio selection, option pricing, hedging and risk management, see e.g. Bauwens, Laurent, and Rombouts (2006). MVM include multivariate GARCH specifications (MGARCH), Multivariate Stochastic Volatility models (MSV) and Multivariate Realized Covariance models (MRC); see McAleer (2005) and Silvennoinen and Teräsvirta (2009) for MGARCH, Asai, McAleer, and Yu (2006) for MSV, and McAleer and Medeiros (2008) for MRC.

A major challenge in MVM is the rapid increase in the number of parameters as the cross-sectional dimension increases. Large cross-sections would be of interest in typical applications of MVM. In unrestricted specifications, however, the number of parameters grows faster than the cross-sectional dimension; this implies that parameters eventually outnumber observations in large cross-sections, a situation where model-estimation becomes infeasible. This phenomenon is similar to the deterioration of the rate of convergence for nonparametric estimators in higher dimensions, see e.g. Linton (2009), and we refer to it as the 'curse of dimensionality' problem for MVM.[1] This curse of dimensionality problem has inspired the use of shrinkage estimators, estimation for sparse covariance matrices and other non-parametric solutions, see e.g. Abadir, Distaso, and Zikes (2010) and reference therein.

In this paper we discuss 'structured' – i.e. restricted – parametric MVM specifications which mitigate or even solve the curse of dimensionality problem. The structure we consider is formalized through weight matrices as in spatial models; in the present context proximity is induced by economic common factors. Spatial models originated as a way to model the joint covariance structure of data coming from different geographical areas, in a single time period; these ideas are applied here for the modeling of conditional covariance matrices over time. Several approaches for the definition of weight matrices are given; the simplest example corresponds to the definition of neighbors as members of the same asset class.

Many empirical studies assume diagonal parameter matrices in MVM. These specifications do not allow for covariance spillover and feedback effects, which are major aspects of interest. On the contrary, structured specifications allow for covariance spillover and feedback effects from neighbors. In this sense structured specifications can reflect the factor structure associated e.g. with the classification of stocks into sectors.

Structured specifications differ from factor volatility models. In the latter, factors are not identified, while in the former they are associated to a precise structure derived from economic rationale.[2] Structured specifications are hence easier to interpret, because factors are defined a-priori.

We discuss the identification of the proposed structured specification; necessary and sufficient conditions for statistical identification are given. Quite obviously, identification has important implications on inference. For unrestricted MGARCH processes, asymptotic properties of unrestricted (quasi-) likelihood-based inference have been discussed in Comte and Lieberman (2003),

[1] Note that for nonparametric estimation the rate of convergence of estimators becomes slower in higher dimensions, while maintaining consistency. For MVM, instead, the curse of dimensionality is more extreme, because estimation simply becomes infeasible in large cross-sections.

[2] In statistical factor analysis, the literature distinguishes between exploratory analysis, where no identifying assumption is made, and confirmatory analysis, where factor loadings are restricted on the basis of a-priori restrictions. Most factor models in MVM are of exploratory nature, and hence leave factor identification unspecified.

Ling and McAleer (2003), Hafner and Preminger (2009b,a), and Francq and Zakoian (2010), inter alia. These papers show that, under appropriate conditions, the (Quasi-) Maximum Likelihood Estimator (QMLE) is Locally Asymptotic Normal (LAN). Here we find that, when the identification conditions are satisfied, the LAN asymptotic results extend directly to the structured specifications. This allows to perform asymptotically χ^2 (robust) Wald and Lagrange Multiplier (LM) tests.

Similarly to spatial econometrics, the specification of weight matrices is based on a-priori knowledge. This knowledge concerns which units are connected, how distances are converted into weights and the pre-determinedness of the weight structure with respect to the data being examined. Different classification criteria may suggest alternative weight matrices, and selecting the right weight matrices hence becomes of practical importance. We indicate that the choice of weight matrices can be addressed within nested model comparisons framework, with asymptotically χ^2 (robust) Wald and LM tests.

Alternatively, models can be compared via tests for equal predictive ability, see e.g. Diebold and Mariano (1995). In order to investigate the finite sample properties of these tests and to evaluate the effects of misspecification, we perform a small Monte Carlo study. Both misspecification of the weight matrix and of homogeneity restrictions on the vector of parameters are considered. We find that tests of predictive ability help to uncover both types of misspecification when the Data Generating Process (DGP) is sufficiently far from the entertained model. When the DGP is not too far from the entertained model, however, parsimonious (but misspecified) models are generally preferred over more flexible (and correctly specified) parameterizations.

The curse of dimensionality problem for MVM is well recognized in the literature. Silvennoinen and Teräsvirta (2009), for instance, provide a list of desiderata for an ideal MGARCH specification, which applies to MVM more in general. The ideal specification should be: i) flexible enough to allow for covariance spillover and feedback effects; ii) scalable, i.e. estimation should be feasible for increasingly large cross-sections; iii) interpretable; iv) easy to estimate numerically. In particular the last property requires that the calculation of the inverse and of the determinant of the conditional covariance matrix should be fast and numerically stable.

We show that proximity structured specifications satisfy most – if not all – these requirements. In fact, most of the structured specifications allow for spillover and feedback effects, i.e. they are flexible. They also have a moderate number of parameters, which usually grows linearly with the cross-sectional dimension; this makes these models scalable. Thanks to the fact that proximity is derived from economic rationale, the corresponding parameters have a direct economic interpretation. Finally, some structured specifications simplify the calculation of the inverse and of the determinant of the conditional covariance matrix with respect to the unrestricted case; in some cases this can lead to a substantial speed up of likelihood computations.

Spatial statistics has a long tradition in modeling unconditional variance matrices, see e.g. Cressie (1993) and references therein. Spatial econometrics has steadily developed over the years, see e.g. Anselin (1988) and reference therein. Applications of spatial models in economics can be found in the literature of regional income growth, the microeconomics of product diffusion, the term structure of interest rates, the analysis of interaction of policy makers in public economics. They are used to account for error dependence in regression models, see e.g. Martellosio (2010), as well as to model individual heterogeneity in panels, see Baltagi, Bresson, and Pirotte (2007), McAleer, Medeiros, and Slottje (2008) and Pesaran and Tosetti (2011). Despite the wide use of spatial models

for unconditional covariance structures, their use in modeling conditional covariances discussed in this paper is novel.[3]

Spatial models are associated with the notion of distance between units. The concept of proximity originated as a geographic concept; coming into economics, it has been associated with economic distance and social proximity, see Conley and Dupor (2003), Pesaran, Schuermann, and Weiner (2004). The simplest notion of proximity is the one inherited from regular lattice models. Case (1991), for instance, classified units into groups, considering units as neighbors if they belong to the same group. We apply this notion to an example of n stock returns, where neighbors are defined as stocks from the same sector. The definition of neighbors is then extended to more general situations, also allowing for the presence of covariates.

The rest of the paper is organized as follows. Section 2 gives the basic setup and the motivation. Section 3 described the structured specifications proposed in this paper, which depend linearly on weight matrices; the latter are discussed in Section 4. Section 5 discusses identification and asymptotics. Section 6 presents a specification analysis of weight matrices and a Monte Carlo study on the effects of misspecification. Section 7 reports an application to six stock returns from the New York Stock Exchange (NYSE) and Section 8 reports concluding remarks. Appendix A discusses properties of proximity matrices, while Appendix B reports proofs on identification.

A final word on notation: 1(condition) denotes the indicator function that takes value 1 if the condition is true and value 0 otherwise; 1_s is an $s \times 1$ vector of ones; vec is the column-stacking operator; $\otimes$ indicates Kronecker product, see Magnus and Neudecker (2007); $\operatorname{diag}(a)$ is the diagonal matrix with vector a on the main diagonal, while $\operatorname{diag}(A_1, \ldots, A_n)$ indicates a block-diagonal matrix with $A_1, \ldots, A_n$ on the main diagonal; $\operatorname{dg}(A)$ indicates the column vector containing the diagonal elements of A; $\operatorname{corr}(A) := \operatorname{diag}(\operatorname{dg} A)^{-1/2} A \operatorname{diag}(\operatorname{dg} A)^{-1/2}$ for a positive definite (p.d.) matrix A; $(A)_{ij}$ indicates element i, j of matrix A; $A_\perp$ indicates a basis of the orthogonal complement of the column space of A.

2. Motivation

We consider a cross-section of n time series $y_t := (y_{1,t} : \cdots : y_{n,t})'$ over a time span $t = 1, \ldots, T$. For instance, y_t may represent n stock returns at time t. Let $\mathcal{F}_t$ indicate the information set up to and including time t, which is generated by the random variables in $z_t := (y_t' : x_t')'$ where x_t contains additional random variables that are observed at time t. Let the conditional mean of y_t be some parametric function of z_{t-1}, $\mu_t(z_{t-1}) := \mathbb{E}_{t-1}(y_t) := \mathbb{E}(y_t|\mathcal{F}_{t-1})$. We concentrate attention to deviations from the conditional mean $u_t := y_t - \mu_t(z_{t-1})$, with $\mathbb{E}_{t-1}(u_t) = 0$. Our interest lies in the prediction of $\Sigma_t = \mathbb{E}_{t-1}(u_t u_t')$. We assume that $u_t = \Sigma_t^{1/2} \varepsilon_t$ where ε_t is i.i.d. with 0 mean and variance matrix equal to I_n. The (quasi) Gaussian log-likelihood function is given by $\ell(\theta) := \log L(\theta) = \sum_{t=1}^T \ell_t(\theta)$, with $\ell_t(\theta) := -\frac{1}{2}(\log\det\Sigma_t + u_t'\Sigma_t^{-1}u_t)$, and θ is the column vector of model parameters.

[3]The authors experimented with selected structured specification in Caporin and Paruolo (2005a,b). After completing the first draft of the present paper in 2008, the authors became aware also of the Dynamic Equicorrelation model (DECO) proposed in Engle and Kelly (2012), which can also be interpreted as a special case of a structured specification.

name	v_t	ξ_t	number of parameters unrestricted	structured
VEC	$\operatorname{vech}\Sigma_t$	$\operatorname{vech} u_t u_t'$	$O(n^4)$	$O(n^2)$
BEKK	$\operatorname{vech}\Sigma_t$	$\operatorname{vech} u_t u_t'$	$O(n^2)$	$O(n)$
GO-GARCH	$\operatorname{dg}(X^{-1}\Sigma_t X^{-1\prime})$	$X^{-1}u_t \odot X^{-1}u_t$	$O(n^2)$	$O(n)$
CCC	$\operatorname{dg}(\Sigma_t)$	$u_t \odot u_t$	$O(n^2)$	$O(n)$
gen. cDCC	$\begin{cases} v_{1t} := \operatorname{dg}\Sigma_t \\ v_{2t} := \operatorname{vech} Q_t \end{cases}$	$\begin{cases} u_t \odot u_t \\ \operatorname{vech} u_t^\dagger u_t^{\dagger\prime} \end{cases}$	$O(n^2)$	$O(n)$
VCC	$\begin{cases} v_{1t} := \operatorname{dg}\Sigma_t \\ v_{2t} := \operatorname{vech} R_t \end{cases}$	$\begin{cases} u_t \odot u_t \\ \operatorname{vech}(\operatorname{corr}(M_{zz}^{(t,h)})) \end{cases}$	$O(n^2)$	$O(n)$
SV	$\log\big(\operatorname{dg}(X^{-1}\Sigma_t X^{-1\prime})\big)$	η_t	$O(n^2)$	$O(n)$
MRV	$\operatorname{vech} Y_t$	ζ_t	$O(n^2)$	$O(n)$

TABLE 1. Examples of the dynamic equation $v_t = \widetilde{c} + \widetilde{A}\xi_{t-1} + \widetilde{B}v_{t-1}$, see eq. (1). $\odot$ is element-wise multiplication. gen. cDCC and VCC present two dynamic equations, one for v_{1t} and one for v_{2t}. gen. cDCC: $u_t^\dagger := \operatorname{diag}(\operatorname{dg} Q_t)^{1/2}\operatorname{diag}(v_{1t})^{-1/2}u_t$, $\Sigma_t = \operatorname{diag}(v_{1t}^{1/2})\operatorname{corr}(Q_t)\operatorname{diag}(v_{1t}^{1/2})$. VCC: $\Sigma_t = \operatorname{diag}(v_{1t}^{1/2})R_t\operatorname{diag}(v_{1t}^{1/2})$, $M_{zz}^{(t,h)} := h^{-1}\sum_{j=0}^{h-1} z_{t-j}z_{t-j}'$, $z_t := \operatorname{diag}(v_{1t})^{-1/2}u_t$. SV, stochastic volatility: η_t is independent of ε_t. MRV: Gourieroux, Jasiak, and Sufana (2009) state directly the transition probability of the Markov process Y_t, which defines implicitly the mean-zero innovation ζ_t.

Many volatility models present a dynamic equation for some vector v_t containing (functions of) variances and covariances in Σ_t of the type

$$v_t = \widetilde{c} + \widetilde{A}\xi_{t-1} + \widetilde{B}v_{t-1} \tag{1}$$

where ξ_t is either a function of u_t or of some random vector η_t independent of ε_t. Table 1 reports several special cases, which include many MGARCH, MSV and MRV models.

As a representative example, consider the following BEKK specification, see Engle and Kroner (1995):

$$\Sigma_t = C + Au_{t-1}u_{t-1}'A' + B\Sigma_{t-1}B', \tag{2}$$

where A, B and C are unrestricted $n \times n$ parameter matrices to be estimated, with C p.d.[4] Taking vecs and using properties of vec and vech operators, one finds that (2) is a special case of (1) with $v_t := \operatorname{vech}\Sigma_t$ and $\xi_t := \operatorname{vech} u_t u_t'$.

The total number of coefficients in (2) is $\frac{1}{2}n(n+1)+2n^2 = O(n^2)$, where $O(n^\alpha)$ indicates the order in terms of n, the cross-sectional dimension. Similar reasoning, applied to the models in Table 1, leads to the numbers of parameters listed in the 4th column of Table 1, labeled 'unrestricted number of parameters'. Columns b and e in Table 2 illustrate the growth in the number of parameters, assuming $T = 1000$ and varying n between 10 and 2000. Observe that estimation becomes infeasible

[4]See also McAleer, Chan, Hoti, and Lieberman (2008) for a motivation of the BEKK specification based on random coefficient autoregressions.

	a	b	c	d	e	f	g
n	$1000n$	$2.5n^2$	$0.5n^4$	$6n$	b/a	c/a	d/a
10	$1{\cdot}10^4$	250	$5{\cdot}10^3$	60	0.025	0.5	0.006
50	$5{\cdot}10^4$	6250	$3125{\cdot}10^3$	300	0.125	62.5	0.006
100	$1{\cdot}10^5$	$25{\cdot}10^3$	$5{\cdot}10^7$	600	0.25	500	0.006
500	$5{\cdot}10^5$	$625{\cdot}10^3$	$3125{\cdot}10^7$	$3{\cdot}10^3$	1.25	62500	0.006
1000	$1{\cdot}10^6$	$25{\cdot}10^5$	$5{\cdot}10^{11}$	$6{\cdot}10^3$	2.5	$5{\cdot}10^5$	0.006
2000	$2{\cdot}10^6$	$1{\cdot}10^7$	$8{\cdot}10^{12}$	$12{\cdot}10^3$	5	$4{\cdot}10^6$	0.006

TABLE 2. Ratio of number of parameters to observations. Entries report number of observations, number of parameters, or their ratio. n: cross-sectional dimension; column a: number of data points for time series with $T = 1000$; columns b, c, d: number of parameters of orders $O(n^2)$, $O(n^4)$, $O(n)$; columns e, f, g: ratio of number of parameters to sample size for columns b, c, d; if > 1, the model is not estimable (more parameters than observations).

between 100 and 500; this illustrates the curse of dimensionality problem. Other columns in Table 2 illustrate the growth in the number of parameters in other models in Table 1; note that the curse of dimensionality applies to any model with order $\alpha > 1$.

3. Structured specifications

This section presents the structured specifications proposed in this paper. Structured specifications make use of weight matrices; a weight matrix W is a known square $n \times n$ matrix whose i, j-th entry w_{ij} indicates the weight (a real number between 0 and 1) of variable j in the determination of variable i. The diagonal entries are equal to 0 and the row-sums may be normalized to be equal to 1. In the following, when different weight matrices are considered, we indicate them as $W^{(h)}$, $h = 0, 1, \ldots$; more details on the definition and the specification of weight matrices are given Section 4. In this section we define structured specifications assuming that at least one weight matrix W is available.

3.1. **Definition of structured specifications.** Consider the example of a BEKK model, see (2), and of a single weight matrix W. A structured specification is obtained by setting $C = S^{-1}VS^{-1\prime}$ and assuming A, B and S to be the following linear functions of the weight matrix W:

$$A = A_0 + A_1 W, \qquad B = B_0 + B_1 W, \qquad S = I - S_1 W, \tag{3}$$

where: $A_j := \mathrm{diag}(\alpha^{(j)})$, $B_j := \mathrm{diag}(\beta^{(j)})$, $j = 0, 1$, $S_1 := \mathrm{diag}(s^{(1)})$, $V := \mathrm{diag}(v)$ are all diagonal $n \times n$ matrices; and $s^{(1)}$, v, $\alpha^{(j)}$, $\beta^{(j)}$, $j = 0, 1$ are $n \times 1$ parameter vectors.

The matrices A, B, S in (3) are special cases of a *proximity matrix*, which we define as any matrix of the form

$$\Pi = \sum_{h=0}^{k} \mathrm{diag}(\psi^{(h)}) W^{(h)}, \tag{4}$$

with $W^{(0)} = I$, and where $\psi^{(h)}$ are $n \times 1$ vectors of coefficients $h = 0, \ldots, p$. Proximity matrices and their properties are discussed in more detail in Appendix A.

In the following $C = S^{-1}VS^{-1\prime}$ with S a proximity matrix, see (4), and V a diagonal matrix will be taken to be the leading example of structured specification for a positive definite matrix. For the

generic parameter matrices $\widetilde{A}$, $\widetilde{B}$ in dynamic equations of the type (1), structured specifications are taken to be proximity matrices, as in (4). In spatial econometrics, row-normalized and unnormalized weight matrices are characterized by different behavior. This is not the case in (4), because the row-normalization can be absorbed in the parameter vectors[5] $\psi^{(h)}$.

3.2. **Number of parameters.** Note that the number of parameters in the structured BEKK specification (3) is $6n = O(n)$, which grows linearly with n and is estimable also for large cross-sections, see columns d and g in Table 2. It can be seen that the reduction in the order $O(n^\alpha)$ is from $\alpha = 4$ or 2 for unrestricted specifications to $\alpha = 2$ or 1 for structured ones. In case $\alpha = 2$ for the structured specification, the latter only mitigates the curse of dimensionality. When instead $\alpha = 1$ for the structured specification, the curse of dimensionality is solved, as the corresponding structured specifications is scalable.

The structured specification $C = S^{-1}VS^{-1\prime}$ ensures that C is p.d. provided $V := \mathrm{diag}(v)$ has positive elements v_i on the main diagonal and S is invertible; the latter is the case provided the diagonal elements in $S_1 := \mathrm{diag}(s^{(1)})$ are different from the reciprocals of the nonzero eigenvalues of W. The implied correlation structure can be quite articulate, despite being governed just by a few coefficients. For example, the specification $S^{-1}VS^{-1\prime}$ for the covariance matrix of a vector u, say, can generate negative correlations between pairs of u_i. This can be seen, for instance, in the simple case of $S = I_2 - \phi W$ with $W = (e_2 : e_1)$, $V = I_2$, where e_i is the i-th column of I_2; the correlation ρ between u_1 and u_2 is $2\phi/(1+\phi^2)$, and one has $-1 < \rho < 1$ for $-1 < \phi < 1$.

The form $C = S^{-1}VS^{-1\prime}$ mimics the form of the variance-covariance matrix of a SAR process, see Cressie (1993), which posits $(I - \phi W)u = \varepsilon$ with scalar ϕ, and errors ε with mean 0 and diagonal covariance matrix $V := \mathrm{diag}(v)$. Provided $S := I - \phi W$ is invertible, one can solve the equations for u by computing $\varepsilon = Su$, from which one finds the SAR covariance structure $\mathbb{E}(uu') = S^{-1}VS^{-1\prime}$.

3.3. **Interpretation.** Despite a moderate number of parameters, and unlike the diagonal BEKK, the structured BEKK specification allows for covariance spillover effects. In fact, consider the term $Au_{t-1}u'_{t-1}A'$ in (2), where

$$Au_{t-1} = \mathrm{diag}(\alpha^{(0)})u_{t-1} + \mathrm{diag}(\alpha^{(1)})Wu_{t-1}.$$

Observe that $(Au_{t-1})_i$ contains two terms: the first one $(\alpha_i^{(0)}u_{i,t-1})$ consists of the own-lag u_{it} term, while the second one is $\alpha_i^{(1)}w_i'u_{t-1}$, a term that delivers the spatial effect from first order neighbors. Here w_i' is the i-th row of W_n and $w_i'u_{t-1}$ is proportional to the average of u_{t-1} for stocks in the same sector; hence $w_i'u_{i,t-1}$ represents spillover effects from other stocks in the sector of unit i. Thus the term $Au_{t-1}u'_{t-1}A'$ contains both diagonal effects and spillover effects from the same sector.

A similar interpretation applies to the $B\Sigma_{t-1}B'$ term in (2). In fact, let e_i be the i-th column of I_n; then $(\Sigma_t)_{ij}$ depends on

$$\left(B\Sigma_{t-1}B'\right)_{ij} = \beta_i^{(0)}\beta_j^{(0)}\left(\Sigma_{t-1}\right)_{ij} + \beta_j^{(0)}\beta_i^{(1)}\left(w_i'\Sigma_{t-1}e_j\right) + \beta_i^{(0)}\beta_j^{(1)}\left(e_i'\Sigma_{t-1}w_j\right) + \beta_j^{(1)}\beta_i^{(1)}\left(w_i'\Sigma_{t-1}w_j\right).$$

The first term contains the own lagged term of the conditional covariance $(\Sigma_{t-1})_{ij}$; the second term contains $w_i'\Sigma_{t-1}e_j$ which is the conditional covariance of the average from neighbors of unit i (excluding asset i) with asset j at time $t-1$; the third term is similar to the second one, interchanging i and j and finally $w_i'\Sigma_{t-1}w_j$ is the conditional covariance of the average neighbors to units i and j

[5] In case $\psi^{(h)}$ is restricted to have all equal elements, $\psi^{(h)} = \psi_h 1_n$, then row-normalization becomes relevant.

(excluding asset i and j). This breakdown clarifies covariance feedback from assets in neighbors of units i and j onto $(\Sigma_t)_{ij}$; the first term represents a diagonal effect, and the last three are feedback effects from the covariances of stocks i and j with their neighbors. This shows the flexibility of the structured specification.

3.4. **Homogeneous and heterogeneous specifications.** The coefficients $\psi^{(h)}$ that appear in (4) are not necessarily distinct. If they are, we call (4) an 'heterogeneous' specification; otherwise, we call it 'homogeneous'. We also consider the case in which $\psi_i^{(h)}$ is equal to $\psi_j^{(h)}$ for all j that are neighbors to unit i with respect to $W^{(h)}$. We call the corresponding specification the 'group-homogeneous' case. Note that the heterogeneous specification nests the group-homogeneous one, which in turn nests the homogeneous one.

Restricted structured specifications can be obtained by considering the group-homogeneous specification, the homogeneous specification or zero-restrictions on a subset of the parameters in $\psi^{(h)}$. Obviously, many sub-models can be constructed by combining restrictions of this type.

3.5. **Computational gains.** In this subsection we illustrate the possible gains in computational speed involved by structured specifications over un-structured ones. We illustrate this property by considering a standard CCC specification. In the CCC model, one specifies $\Sigma_t = D_t R D_t$, where R is a correlation matrix, $D_t = \mathrm{diag}(h_t)$ and $v_t := h_t \odot h_t$ satisfies the GARCH equation (1) with $\xi = u_t \odot u_t$.

Evaluation of the log-likelihood requires the computation of $\Sigma_t^{-1} = D_t^{-1} R^{-1} D_t^{-1}$ and of $\log\det\Sigma_t = \sum_{i=1}^n \log v_{it} + \log\det R$. For the standard, un-structured specification, this requires the computation of $\log\det R$ and of R^{-1} using standards algorithms.

A possible structured specification is $R = \mathrm{corr}(C) = G^{-1} C G^{-1}$ with $C = S^{-1} V S'^{-1}$, $S = I - \phi W$, $V = \psi I$, $G := \mathrm{diag}(\mathrm{dg}(C))^{1/2}$. For concreteness, we also specify W as $J_n = (n-1)^{-1}(1_n 1_n' - I)$. In the structured specification, the computation of the inverse R^{-1} simplifies as follows:

$$R^{-1} = GC^{-1}G = GSV^{-1}S'G = \frac{1}{\psi} G\,(I - \phi W)\,(I - \phi W)'\,G, \tag{5}$$

which does not require the use of inversion routines. In the calculation of $\log\det\Sigma_t$, one finds

$$\log\det R = -\sum_{i=1}^{n} \log C_{ii} - 2\log\det S + n\log\psi \tag{6}$$

Expression (6) does not involve simplifications for generic W, given that the r.h.s. contains $\log\det S$. For the choice $W = J_n$, however, the expression of the determinant of S can be simplified; in fact, let $P := n^{-1} 1_n 1_n'$, and note that one can re-write S as a linear combinations of P and $I - P$, $S = I - \frac{\phi}{n-1}(1_n 1_n' - I) = (1-\phi)P + (1 + \frac{\phi}{n-1})(I - P)$. Because P and $I - P$ are orthogonal projectors, one can apply Lemma 2.1-iii in Magnus (1982) to find

$$\log\det S = \log(1-\phi) + (n-1)\log\left(1 + \frac{\phi}{n-1}\right) \tag{7}$$

which involves just a scalar computation.

Table 3 reports the ratio between the average computing time of R^{-1} (respectively $\log\det R$) using generic matrix routines and the average computing time using eq. (5),(respectively (6) and (7)) for the structured specifications.[6] For the computation of R^{-1} in the structured case of eq.

[6]We used version 7.9.0 of Matlab 64 bit for Windows, on a desktop personal computer with Intel processor EM64T and 12Gb of RAM.

n	10	50	100	500	1000
R^{-1}	3.44	5.76	4.31	1.72	1.53
$\log\det R$	0.84	7.31	22.87	236.94	715.87

TABLE 3. Ratio between the mean computing times using generic matrix routines and using eq.s (5), (6) and (7). R^{-1}: ratio between mean computing times using Matlab functions `inv` and using eq. (5). $\log\det R$: ratio between the mean computing times using Matlab function `det` and using eq. (6) and (7). Means were calculated over 200 replications.

	$G=1$	2	3
$F=1$	1,2	3,4	5,6
2	7,8	9,10	11,12

TABLE 4. Example of a two-way classification of 12 assets, see Section 4. Entries are labels of 12 stocks classified by the two factors F and G.

(5) we used the fact that in this model the diagonal elements of G are all equal. Table 3 shows that the computational gains can be substantial. Interestingly, the computational gains seem to be decreasing in n for R^{-1} and increasing in n for $\log\det R$. The actual magnitude of the computational gains is, in general, hardware- and software-dependent.

4. Weight matrices

Any proximity-structured specification requires the availability of one or more weight matrices $W^{(h)}$. In this section we discuss the definition of weight matrices in the context of MVM, where weight matrices have dimension n when they describe proximity of vector u_t, and have dimension n^2 when they describe proximity of $\mathrm{vec}\Sigma_t$. We also cover the more general case of weight matrices that reflect a (possibly time-dependent) metric distance between units. The econometric specification analysis of weight matrices is deferred to Section 6.

4.1. **One classification criterion.** Assume that there are $n = 12$ assets returns, classified on the basis of each stock's sector, represented by the factor F with levels 1 (goods sector) and 2 (service sector). Stocks are labeled according to their position in the u_t vector, and we assume that the classification of stock labels detailed in Table 4 applies.

Next define a weight matrix $W^* := (w^*_{ij})$ which classifies asset returns i and j as neighbors if the two stocks belong to the same sector, $F_i = F_j$

$$w^*_{ij} := 1\left(F_i = F_j, i \neq j\right) \tag{8}$$

In the example of Table 4 one finds that $W^* = 5\,\mathrm{diag}(J_6, J_6)$ where $J_s := (s-1)^{-1}(1_s 1_s' - I_s)$.

Each row i in a weight matrix W^* represents the weights of the neighbors of unit i; the number of neighbors may differ across rows. One can normalize W^* for the different row-sums by applying the transformation

$$w_{ij} := \begin{cases} w^*_{ij} / \sum_{j=1}^n w^*_{ij} & \text{if } \sum_{j=1}^n w^*_{ij} > 0 \\ 0 & \text{otherwise} \end{cases}, \tag{9}$$

which produces a normalized weight matrix $W = (w_{ij})$. Because the matrix W reflects the classification criterion F, in the following we indicate it by W^F. In the case of the example, $W^F = \text{diag}(J_6, J_6)$, because all the rows of W^* have the same row-sums.

4.2. **Several classification criteria.** When more than one classification criterion is present, one can construct weight matrices applying the same principles and techniques in use in Analysis of Variance, ANOVA, see e.g. Wichura (2006).

In the example of the 12 assets in Table 4, consider a second classification criterion G representing capitalization size, with levels 1, 2, 3 corresponding to 'large', 'medium' and 'small'. Let W^G be the weight matrix constructed in the same way as W^F when using classification criterion G in place of F. One possible combination of the factors F and G can be obtained by considering a proximity matrix $S = \text{diag}(\psi_F)W^F + \text{diag}(\psi_G)W^G$; here the effects of both factors are additive.

A more general combination of factors F and G can be obtained by considering each cell in Table 4 as level of the factor $H := F \times G$; this allows to measure interactions between the two factors. Specifically the level h of the factor H corresponds to the pair (i, j) for (F, G), with $h = (k_1 - 1)i + j$, where k_1 (k_2) is the number of distinct values of F (G). The combined factor H presents $k_1 \cdot k_2$ intensities, and one can define a weight matrix corresponding to it, labeled W^H, as detailed above for factor F. Specifically, for the example in Table 4 one finds $W^H = \text{diag}(J_2, J_2, J_2, J_2, J_2, J_2)$. This can be directly extended to the case of several factors, where we note that more refined classifications obviously imply fewer units per group. This gives an implicit upper limit to the number of factors one can consider in practice.

4.3. **From stocks weights to covariances weights.** The previous subsections have shown how the classification of stocks gives rise to weight matrices W_n for stock returns, where in this subsection the subscript n to the weight matrix indicates its dimension. Here we show how these matrices can then be used to derive weight matrices for stock returns (co-)variances.

Recall that the generic weight matrix for a single classification criterion is given by

$$W_n := \text{diag}(J_{n_1}, \ldots, J_{n_{k_1}}), \tag{10}$$

where the n_i stocks if class i are ordered consecutively for $i = 1, \ldots, k_1$.

When considering $\text{vec}(\Sigma_t)$ or $\text{vec}(u_t u_t')$, element number h in these vectors corresponds to a pair of row- and column- indices (i, j), where $h = (i-1)n + j$; let also (l, m) correspond to element $v = (l-1)n + m$. In order to construct weight matrices for the elements in $\text{vec}(\Sigma_t)$, one may define the un-normalized weight as

$$w^*_{hv} := 1(F_i = F_l, F_j = F_m, i \neq l, j \neq m), \tag{11}$$

where we have used an expression similar to (8). Alternative weight matrices are obtained by replacing w^*_{hv} in (11) with either one of the following expressions:

$$w^{*(1)}_{hv} := 1\left(F_i = F_l, i \neq l, j = m\right), \qquad w^{*(2)}_{hv} := 1\left(F_j = F_m, i = l, j \neq m\right). \tag{12}$$

In words, (11) states that the pairs (i, j) and (l, m) are neighbors if sector pairs are the same and no stock appears twice. The two weights in (12) require instead that either one of the two stocks is the same. $w^{*(1)}_{hv}$ requires that the second stock in the pairs to be the same, while $w^{*(2)}_{hv}$ requires that the first stock in the pairs to be the same; the remaining stock in the pair needs to be different.

Note that the three cases in (11), (12) are mutually exclusive. One can then prove the following theorem.

Theorem 1 (Weights in n and n^2 dimensions)**.** *Let $W_{n^2}^{(j)}$ be the weight matrix obtained by normalizing the weights $w_{hv}^{*(j)}$ in* (12)*, $j = 1, 2$ and let $W_{n^2}^{(3)}$ be the weight matrix obtained by normalizing the weights in eq.* (11)*. Finally let W_n be the weight matrix for stocks returns, as in* (10)*. Then one has*

$$W_{n^2}^{(1)} = I_n \otimes W_n, \qquad W_{n^2}^{(2)} = W_n \otimes I_n, \qquad W_{n^2}^{(3)} = W_n \otimes W_n.$$

In general the Kronecker product $H \otimes K$ is a weight matrix provided either H or K is a weight matrix and the other matrix has non-negative entries with row sums equal to 1*; this holds in particular when H and K are both (normalized) weight matrices.*

The first part of Theorem 1 can be applied to find the nesting relation between the structured-BEKK and the structured-VEC specification. Assume that the matrices A, B in (2) satisfy the restrictions (3). Taking vecs of equation (2), one finds

$$\text{vec } \Sigma_t = \text{vec } C + (A \otimes A) \text{ vec } \left(u_{t-1} u_{t-1}'\right) + (B \otimes B) \text{ vec } \Sigma_{t-1},$$

where we have used standard properties of vecs. The matrices $A \otimes A$ and $B \otimes B$ are spatial matrices of the type (4) with respect to the set of weight matrices $\mathbb{W} := \{W_{n^2}^{(j)}, j = 1, 2, 3\}$ by Theorem 1. We have hence proved the following corollary.

Corollary 2 (s-BEKK is nested within s-VEC)**.** *The structured-BEKK specification is nested within the structured-VEC MGARCH specification.*

4.4. **Weight matrices using covariates.** Weight matrices can be defined using covariates x_{t-1} (which are measurable with respect to $\mathcal{F}_{t-1}$) into a weight matrix W_t, such as market value, book-value, momentum, earnings/price, cash-flow/price, dividend yield, short- and long-term reversals. Another set of covariates can be constructed as dissimilarity between sectors as represented by appropriate columns in input-output matrices, as in Conley and Dupor (2003).

Let $x_{i,t}$ be a $q \times 1$ vector of indicators available at time t concerning assets i, which are measurable with respect to $\mathcal{F}_t$. For simplicity we assume that each entry in $x_{i,t}$ is non-negative and it is normalized to be on a scale from 0 to 1. Next define the un-normalized weight matrix $W_t^* := (w_{i,j,t}^*)$ with weights

$$w_{i,j,t}^* := (1 - \delta_{i,j}) \exp\left(-s \left\| x_{i,t-1} - x_{j,t-1} \right\|_b^r\right), \qquad b, r, s > 0, \tag{13}$$

where $\delta_{i,j}$ is Kronecker's index, b, r and s are positive constants and $\|a\|_b := \left(\sum_{i=1}^q |a_i|^b\right)^{1/b}$. The normalized weight matrix $W_t := (w_{i,j,t})$ is obtained by row-normalization of W_t^*, as in (9).

Note that this definition of the weight matrix reduces to the one associated with a classification criterion F when $x_{i,t} := F_i$, for any choice of $s > 0$, $r \geq 0$. Hence (13) provides a generalization on how to define weight matrices. The choice of the exponential $\exp(-sx^r)$ in (13) has the disadvantage of producing positive real numbers for $w_{i,j,t}^*$ for all values of x, even when x is very far from 0. This implies that each row in W is full. In order to obtain more sparse weight matrices, one can replace $\exp(-sx^r)$ with $\exp(-sx^r)1(x < c)$, that sets all weights equal to 0 for $x \geq c$.

This discussion shows that any discrete or continuous, time-varying or time-invariant covariate can be used to define weight matrices. In the rest of the paper we restrict attention to the time-invariant case.

5. Identification and asymptotics

In this section we discuss identification and asymptotics of structured specifications, where identification is understood in the sense of Rothenberg (1971). We first discuss identification conditions for dynamic parameters A and B, and next we treat the positive definite specification $C = S^{-1}VS^{-1\prime}$. Finally, we present (robust) Wald and LM tests of the restrictions associated with structured specifications when identification holds.

5.1. **Dynamic parameter matrices.** Let Π indicate either matrix A or matrix B in the dynamic equation (2), and assume that Π satisfies the proximity matrix specification in (4). Here Π is $n \times n$ and ψ denotes the associated parameter vector in (4), with $\psi := (\psi_1' : \cdots : \psi_n')'$, $\psi_i := (\psi_i^{(0)} : \cdots : \psi_i^{(k)})'$. With this definition, one finds $\mathrm{vec}(\Pi') = M\psi$ where $M := \mathrm{diag}(M_1, \ldots, M_n)$ is $n^2 \times n(k+1)$, $M_i := (W_i^{(0)} : \cdots : W_i^{(k)})$ and $W_i^{(h)\prime}$ is the i-th row of $W^{(h)}$.

The parameter ψ corresponds to an heterogeneous specification; the group-homogeneous or the homogeneous specifications correspond to linear restrictions on ψ of the generic type $\psi = H\varphi$, where H is a full column rank matrix. Indicate by n_ψ, n_φ the dimensions of the parameter vectors ψ and φ respectively, and partition the parameter vector in two parts, ψ and η.

Let $\ell(\psi, \eta)$ be the log-likelihood of the model, or the function to be optimized in model estimation. We say that ψ is identified if $\psi_1 \neq \psi_2$ implies $\ell(\psi_1, \cdot) \neq \ell(\psi_2, \cdot)$ with positive probability. If this property holds only locally in an open set in the parameter space which includes the true value, we say that ψ is locally identified. We use a similar terminology also in relation to the unrestricted model, i.e. the model where Π is not restricted by (4).

Theorem 3 (Identification of dynamic parameter matrices)**.** *Let Π be globally (locally) identified in the unrestricted model, where Π is either A or B, and let ψ be the parameters of the structured specification; then a necessary and sufficient condition for the global (local) identification of ψ is*

$$M \text{ is of full column rank } n_\psi, \tag{14}$$

where $M := \mathrm{diag}(M_1, \ldots, M_n)$, $M_i := (W_i^{(0)} : \cdots : W_i^{(k)})$ and $W_i^{(h)\prime}$ is the i-th row of $W^{(h)}$.

Similarly, a necessary and sufficient condition for the global (local) identification of the restricted structured parameter vector φ is

$$MH \text{ is of full column rank } n_\varphi. \tag{15}$$

Note that for MGARCH models, only local identification applies. Remark that the matrices in the rank conditions (14), (15) do not involve the parameters, but they are functions of the constraints alone.

The identification conditions (14) or (15) are necessary and sufficient, i.e. they are 'rank conditions'. The corresponding 'order conditions' are found by requiring that the number of rows in W in (14) and MH in (15) are greater or equal to the number of columns; this proves the following corollary.

Corollary 4 (Number of weight matrices)**.** *A necessary but not sufficient condition (order condition) for the heterogeneous specification parameters ψ to be identified is*

$$k \leq n - 1. \tag{16}$$

For the restricted structured parameter φ, a necessary but not sufficient condition (order condition) for identification is

$$n_\varphi \leq n^2. \tag{17}$$

5.2. **Positive definite matrices.** Consider the case where a symmetric and positive definite $n \times n$ matrix C is identified in the unrestricted MVM model. We consider the structured specification $C = S^{-1}VS^{-1\prime}$ with

$$S := I + \sum_{h=1}^{k} \operatorname{diag}(\psi^{(h)})W^{(h)}, \qquad V = \operatorname{diag}(v), \tag{18}$$

where $v := \sigma \odot \sigma$, $\sigma := (\sigma_1, \ldots, \sigma_n)'$. Let ψ be the $n_\psi \times 1$ parameter vector for the proximity matrix S, defined similarly to Π above, and $\psi := (\psi_1' : \cdots : \psi_n')'$. $\psi_i := (\psi_i^{(0)} : \cdots : \psi_i^{(k)})'$. Let σ be the $n \times 1$ parameter vector containing the square root of the diagonal elements of matrix V; here $\gamma := (\psi' : \sigma')'$ are the parameters of interest; let also η indicate the remaining parameters. Similarly to above, we define the matrices $M := \operatorname{diag}(M_1, \ldots, M_n)$, $M_i := (W_i^{(1)} : \cdots : W_i^{(k)})$, where $W_i^{(h)\prime}$ is the i-th row of $W^{(h)}$; note that here, however, the matrix $W^{(0)} := I$ does not appear and that hence M is $n^2 \times nk$.

We consider generic linear restrictions of the type

$$\psi = H_\psi \varphi + h_\psi, \qquad \sigma = H_\sigma \zeta + h_\sigma, \tag{19}$$

which result in linear constraints of the type $\gamma = H\xi + h$ with $\xi := (\varphi' : \zeta')'$, $h := (h_\psi' : h_\sigma')'$ and $H = \operatorname{diag}(H_\psi, H_\sigma)$; in other words we require no cross-restrictions between ψ and σ. Let also n_ξ indicate the number of elements in ξ.

In the following we say that a parameter vector is generically locally identified if it is locally identified for all parameter values in an open set containing the true value except for a set of Lebesgue measure 0, see e.g. Lucchetti (2006). Using results in the latter paper, we can prove the following Theorem 5. In the statements below, $\mathcal{K}_n$ represents the commutation matrix of order n and G is the matrix that satisfies $\operatorname{vec}(\operatorname{diag}(a)) = Ga$ for any vector a, see Magnus (1988), Magnus and Neudecker (2007).[7]

Theorem 5 (Identification of positive definite structured matrices)**.** *Let C be a symmetric, positive definite locally (generically) identified parameter in the unrestricted model; then a necessary but not sufficient condition for γ to be locally identified is*

$$2k \leq n - 1. \tag{20}$$

A necessary and sufficient condition for γ to be locally generically identified is that eq. (20) *holds and*

$$\operatorname{rank} \begin{pmatrix} U_\psi & 0 \\ U_\sigma & T_\sigma \widetilde{D}_n \end{pmatrix} = n^2 + n(n-1)/2, \tag{21}$$

where $\widetilde{D}_n$ is a $n^2 \times n(n-1)/2$ matrix whose columns for a basis for space of vectors of the form $\operatorname{vec} A$, where A is any $n \times n$ skew-symmetric matrix[8]*, $q_\psi = \operatorname{vec} I_n$, $q_\sigma = 0$, $N_\psi := \mathcal{K}_n M$, $N_\sigma := G$,*

[7] For instance, one has the representations $G = \sum_{i=1}^n (e_i e_i' \otimes e_i) = \sum_{i=1}^n (e_i \otimes e_i e_i')$.

[8] For instance one can take $\widetilde{D}_n = (D_1 : \cdots : D_{n-1})$ with $D_j := e_j \otimes F_j - F_j \otimes e_j$, $F_j := (e_{j+1} : \cdots : e_n)$.

$n_\psi := nk$, $n_\sigma := n$. *Moreover, for* $a = \psi, \sigma$ *the following notation is employed:*

$$U_a := \begin{pmatrix} R_a' \left(N_{a,1}' \otimes I_n\right) \\ \vdots \\ R_a' \left(N_{a,n_a}' \otimes I_n\right) \\ R_a' \left(N_{a,0}' \otimes I_n\right) \end{pmatrix}, \qquad T_\sigma := \begin{pmatrix} R_\sigma' \left(I_n \otimes N_{\sigma,1}\right) \\ \vdots \\ R_\sigma' \left(I_n \otimes N_{\sigma,n_\sigma}\right) \\ R_\sigma' \left(I_n \otimes N_{\sigma,0}\right) \end{pmatrix},$$

where, for $i = 1, \ldots, n_a$, *the* $n \times n$ *matrix* $N_{a,i}$ *is such that* $\mathrm{vec}(N_{a,i})$ *equals the* i*-th column of* N_a, *while* $N_{a,0}$ *is such that* $\mathrm{vec}(N_{a,0}) := q_a$ *and* $R_a := N_{a\perp}$.

For the restricted parameter vector ξ, *the order condition is replaced by*

$$2n_\xi \leq n(n+1). \tag{22}$$

The rank condition for ξ *to be locally generically identified is that eq.* (21) *holds with* $a = \varphi, \zeta$ *and* $N_\varphi := \mathcal{K}_n M H_\psi$, $N_\zeta := G H_\sigma$, $q_\varphi := \mathrm{vec}\, I_n + \mathcal{K}_n M h_\psi$, $q_\zeta := G h_\sigma$.

We note that the matrix in the rank condition (21) does not involve the parameters, but it is a function of the constraints alone.

5.3. **Asymptotics.** This subsection shows how asymptotic results derived for unrestricted MVM specification extend to structured ones when the conditions of identification hold. For unrestricted MGARCH processes, the LAN property of the QMLE for unstructured BEKK models has been discussed in Comte and Lieberman (2003). Hafner and Preminger (2009a,b) treat the VEC specification and Factor MGARCH. Ling and McAleer (2003) and Francq and Zakoian (2010) cover the CCC specification.

In the rest of this subsection we adopt notation similar to Davidson and MacKinnon (1993) Chapter 8. Let $\ell(\theta) := \sum_{t=1}^T \ell_t(\theta)$, $g(\theta) := \partial \ell(\theta)/\partial \theta'$, $g_t(\theta) := \partial \ell_t(\theta)/\partial \theta$, $G(\theta) := (g_1(\theta) : \cdots : g_T(\theta))'$, $H(\theta) := \partial^2 \ell(\theta)/\partial \theta \partial \theta'$, $H_t(\theta) := \partial^2 \ell_t(\theta)/\partial \theta \partial \theta'$, and indicate the unrestricted and the structured QMLE of the $d \times 1$ vector of parameters θ as $\widehat{\theta}$ and $\widetilde{\theta}$. Assume for simplicity, that z_t is strictly stationary and ergodic. Structured specifications involve smooth (i.e. continuously differentiable) restrictions, which are indicated here in implicit form as $r(\theta) = 0$, where $r(\theta)$ is a $q \times 1$ vector, and in explicit form as $\theta = \theta(\rho)$, where ρ is an $m \times 1$ vector of parameters, $d = m + q$. We also indicate the first order derivatives matrices of the constraints as $R'(a) := \partial r(\theta)/\partial \theta'|_{\theta=a}$ and $Q(b) := \partial \theta(\rho)/\partial \rho'|_{\rho=b}$. Let θ_0 be the true value of θ, and indicate as $g_0 = g(\theta_0)$, $\widehat{g} = g(\widehat{\theta})$, $\widetilde{g} = g(\widetilde{\theta})$ and similarly for other quantities.

Assume that $\widehat{\theta}$ is a consistent solution to the QML problem, and that the classical conditions for LAN of the QMLE are satisfied. These results are usually based on a first order approximation of the score in a neighborhood of θ_0, where $g_{t,0}$ has mean $E(g_{t,0}) = 0$ and nonsingular variance $\mathcal{J}_0 := E(g_{t,0} g_{t,0}') < \infty$, and the expected hessian $\mathcal{H}_0 := E(H_{t,0})$ is nonsingular;[9] this corresponds to the identification of the unstructured MVM. Using the first order approximation of the score, one obtains $T^{1/2}(\widehat{\theta} - \theta_0) \simeq -\mathcal{H}_0^{-1}(T^{-1/2} g_0) \xrightarrow{w} N(0, V)$, with $V := \mathcal{H}_0^{-1} \mathcal{J}_0 \mathcal{H}_0^{-1}$.

Under these conditions and the full column rank of $Q(\theta)$ for θ in a neighborhood of θ_0 (which is implied by the rank conditions for identification), the same consistency and LAN results apply to the structured QMLE estimator $\widetilde{\theta}$, see e.g. Billingsley (1961) Section 3. Specifically, there exists a consistent root $\widetilde{\theta}$ of the constrained structured specification with probability tending to

[9] Here E is with respect to the probability measure indexed by θ_0. In the following $\simeq$ indicates quantities that are equal up to a term that tends to 0 in probability and $\xrightarrow{w}$ indicates weak convergence.

1, and $T^{1/2}(\widetilde{\theta}-\theta_0)$ is asymptotically normal. Again these results are obtained using a first-order approximation for the (restricted) score.

In this framework, one can derive (robust) Wald and LM tests of the restrictions implied by the structured specification, in the spirit of Bollerslev and Wooldridge (1992);[10] this is illustrated in the rest of this section. Wald tests can be constructed by applying the delta-method to the function $r(\theta)$. In fact, one has $T^{1/2}(\widehat{r}-r_0)\simeq -T^{-1/2}R_0'\mathcal{H}_0^{-1}g_0 \xrightarrow{w} N(0,R_0'VR_0)$ and a robust Wald test statistic is given by $T\widehat{r}'(\widehat{R}'\widehat{V}\widehat{R})^{-1}\widehat{r}\xrightarrow{w}\chi^2(q)$, where $\widehat{V}=T\widehat{H}^{-1}\widehat{G}'\widehat{G}\widehat{H}^{-1}$ is a consistent estimate of V, see e.g. Davidson and MacKinnon (1993). Under correct specification of the likelihood, the standard Bartlett identity $\mathcal{H}_0=-\mathcal{J}_0$ holds, $V=\mathcal{J}_0^{-1}$, and also the non-robust Wald test statistics built with $\widehat{V}=T\widehat{H}^{-1}$or $\widehat{V}=T(\widehat{G}'\widehat{G})^{-1}$ are asymptotically $\chi^2(q)$ distributed.

Similarly, one can define robust LM tests as appropriate quadratic forms in $\widetilde{g}$. Standard derivations (see e.g. Davidson and MacKinnon (1993) eq. (8.72)-(8.75)) and the first-order expansion of the score imply that $T^{-1/2}\widetilde{g}\simeq T^{-1/2}R_0F_0'g_0$, $F_0'=(R_0'\mathcal{H}_0^{-1}R_0)^{-1}R_0'\mathcal{H}_0^{-1}$, where $T^{-1/2}g_0\xrightarrow{w}N(0,\mathcal{J}_0)$. Hence $T^{-1/2}\widetilde{F}'\widetilde{g}\simeq T^{-1/2}F_0'g_0\xrightarrow{w}N(0,U)$ with $U=F_0'\mathcal{J}_0F_0$, and a robust LM test statistics can be defined as $T^{-1}\widetilde{g}'\widetilde{F}\widetilde{U}^{-1}\widetilde{F}'\widetilde{g}$, where $\widetilde{U}=T^{-1}\widetilde{F}'\widetilde{G}'\widetilde{G}\widetilde{F}$, with $\widetilde{F}'=(\widetilde{R}'\widetilde{H}^{-1}\widetilde{R})^{-1}\widetilde{R}'\widetilde{H}^{-1}$. Similarly to the Wald test case, under correct specification of the likelihood, the standard Bartlett identity $\mathcal{H}_0=-\mathcal{J}_0$ holds, $U=(R_0'\mathcal{H}_0^{-1}R_0)^{-1}$, and also the non-robust LM test statistics built with $\widetilde{U}=T^{-1}(\widetilde{R}'\widetilde{H}^{-1}\widetilde{R})^{-1}$ or $\widetilde{U}=T^{-1}(\widetilde{R}'(\widetilde{G}'\widetilde{G})^{-1}\widetilde{R})^{-1}$ are asymptotically $\chi^2(q)$ distributed.

Note that in the calculation of the LM test statistics one can use that $R'(\theta(\rho))Q(\rho)=0$ which holds by the chain rule of differentiation. Under identification $\widetilde{Q}$ is of full column rank; hence one can compute $\widetilde{R}$ from $\widetilde{Q}$ as $\widetilde{R}=\widetilde{Q}_\perp$; see Paruolo (1997) on various ways to compute a basis of the orthogonal complement. This observation can be applied, for instance, to the specification (18), which is in explicit form.

6. Specification of weight matrices

In this section we discuss the econometric specification analysis of weight matrices. Two approaches are entertained; the first one employs Wald and LM test for nested model comparisons, using results from the previous section. The second approach is based on tests of equal predictive ability. Finally, we investigate the effects of possible misspecification of the weight matrix on model selection via tests of equal predictive ability through a small Monte Carlo.

6.1. **Likelihood-based specification tests.** In applications, the researcher may entertain the specification of two or more weight matrices W_i. The specification of proximity matrices (4) allows to include several W_i matrices, as long as the identification conditions discussed in Section 5 are satisfied. This allows to nest different specifications of W_i matrices within a single structured proximity matrix. Under identification, the MVM model can be estimated including the weight matrix W^* along with other weight matrices in (4); a Wald test can then be constructed by restricting the parameters ψ^* that multiply W^* within (4) to zero. This procedure delivers Wald-type specification tests on W^* within a given structured specification. An alternative procedure is to first estimate a model with one choice of weight matrix, and then use a Lagrange Multiplier (LM) test to ascertain if an additional weight matrix W^* is needed to express the proximity structure of the model. Under identification, this approach also provides a viable specification analysis procedure.

[10] See also Silvapulle and Silvapulle (1995) Section 2.1 and reference therein.

Hence, when the number of W_i matrices is moderate, one can test which W_i matrices to retain using either Wald or LM tests within a nested hypothesis-testing framework. This contrasts with the specification of weight matrices in the spatial econometrics literature, where this problem has been addressed via the definition of J-type tests for non-nested alternatives, see the spatial J test discussed in Kelejian (2008); see also Burridge and Fingleton (2010) for small-sample alternatives.

6.2. **Tests of equal predictive ability.** An alternative approach is to compare the predictive ability of models with different weight matrices. This is the approach also taken the empirical application of Section 7.

In order to illustrate this approach we designed a small Monte Carlo experiment. We generated data on six variables using a structured BEKK model, see eq. (3). For simplicity, we refer to each simulated variable as an asset return. The conditional expectation of the time series were set to zero. The six variables were specified to belong to two groups of size three, with normalized weight matrix W equal to $W = \mathrm{diag}(J_3, J_3)$. The intercept C was set to the identity matrix, A and B were chosen as proximity matrices which ensure that the unconditional covariance is positive definite. The values of the parameters were chosen as follows:

$$\begin{aligned}
\text{DGP1}\quad & a_0 = (0.06, 0.08, 0.10, 0.26, 0.28, 0.30)',\ a_1 = (0.10, 0.08, 0.06, 0.20, 0.18, 0.16)', \\
& b_0 = (0.74, 0.76, 0.78, 0.44, 0.46, 0.48)',\ b_1 = (0.12, 0.10, 0.08, 0.22, 0.20, 0.18)'; \\
\text{DGP2}\quad & a_0 = (0.06, 0.08, 0.10, 0.06, 0.08, 0.10)',\ a_1 = (0.10, 0.08, 0.06, 0.10, 0.08, 0.06)', \\
& b_0 = (0.84, 0.86, 0.88, 0.84, 0.86, 0.88)',\ b_1 = (0.12, 0.10, 0.08, 0.12, 0.10, 0.08)'.
\end{aligned}$$

In DGP1 the dynamic parameters of the two groups are quite different, while heterogeneity within groups is small. In DGP2 the dynamic parameters of the two groups are identical. Also for DGP2, the degree of heterogeneity within groups is small. We simulated series of 1500 observations using Gaussian innovations, setting $T = 2000$ and discarding the initial 500 observations, to avoid influence from initial values. Using the simulated series, we fitted the following models:

- M1: the correctly specified Structured BEKK model, with an heterogeneous specification;
- M2: a structured BEKK model with homogeneous parameter matrices, using the correct $W = \mathrm{diag}(J_3, J_3)$ matrix;
- M3: a structured BEKK model with heterogeneous parameter matrices, but with misspecification in the groups definition; here asset 4 was erroneously associated with the first group. The (misspecified) weight matrix used in estimation was $W^* := \mathrm{diag}(J_4, J_2)$;
- M4: a structured BEKK model with homogeneous parameter matrices that combines the misspecifications of M2 and M3.

Note that in model M2 the misspecification comes from the restriction of equal parameters for assets belonging to the same group. In model M3, it comes from the wrong classification into groups. Model M4 combines both types of misspecification.

Because the models are not all nested, we chose to evaluate the performance of M1, M2, M3 and M4 using their forecasts. The models were estimated on the first 1250 simulated data, and model forecasts were evaluated on the last 250 observations, using the 250 one-step-ahead covariance forecasts. Models were not re-estimated in the forecast periods.[11]

[11]Calculations in this and the following sections were performed in GAUSS 9. Sample GAUSS programmes for the implementation of a selection of the models in this paper are available at the corresponding author's web page.

	DGP1				DGP2			
Comparison	AG	L_1	L_2	L_3	AG	L_1	L_2	L_3
M1 is better than M2	0.008	0.170	0.131	0.008	0.006	0.025	0.072	0.078
M1 is worse than M2	0.026	0.502	0.746	0.727	0.039	0.864	0.832	0.435
M1 is better than M3	0.735	0.997	0.999	0.206	0.249	0.915	0.933	0.479
M1 is worse than M3	0.000	0.000	0.000	0.007	0.003	0.020	0.018	0.099
M1 is better than M4	0.826	0.998	0.999	0.152	0.322	0.222	0.203	0.299
M1 is worse than M4	0.000	0.000	0.001	0.017	0.415	0.556	0.709	0.174

TABLE 5. Frequency of model selection in bivariate model comparisons. AG: Amisano and Giacomini (2007) test. L_i: Diebold and Mariano (1995) test with loss function L_i in eq. (23), $i = 1, 2, 3$.

Covariance forecasts were compared to the true conditional covariance matrix by means of the loss functions proposed in Laurent, Rombouts, and Violante (2012):

$$L_1 = \operatorname{tr}\left((\Sigma_t - H_t)'(\Sigma_t - H_t)\right), \qquad L_2 = \operatorname{tr}\left(H_t^{-1}\Sigma_t\right) - \log|H_t^{-1}\Sigma_t| - n, \tag{23}$$

$$L_3 = \frac{1}{6}\operatorname{tr}\left(\Sigma_t^3 - H_t^3\right) - \frac{1}{2}\operatorname{tr}\left(H_t^2(\Sigma_t - H_t)\right), \tag{24}$$

where Σ_t denotes the true conditional covariance matrix and H_t is the forecasted conditional covariance. The first loss function L_1 is the Frobenius norm and represents the mean squared error loss for matrices. The second loss function L_2 is the loss of James and Stein (1961) and penalizes under-predictions. Finally, the third loss function L_3 penalizes (co-)variance over-predictions.

We tested equality of predictive ability applying the Diebold-Mariano test, DM, to the loss functions in eq. (23). We also compared models by means of the forecast test proposed in Amisano and Giacomini (2007), AG. The AG test compares forecast performances in terms of a Gaussian likelihood of competing models using the one-step-ahead covariance forecasts produced by the different models.

The results are reported in Table 5. For DGP1, the misclassification of asset 4 is well detected by loss functions L_1 and L_2, and $M1$ is preferred to $M3$ and to $M4$ with high frequency. A similar performance is displayed by the AG test, while the Loss Functions L_3 indicates equivalence between $M1$ and $M3$ or $M4$ in over 75% of cases. The comparison of models $M1$ and $M2$ shows that, given the small heterogeneity within group, the group-homogeneous model $M2$ is preferred to $M1$ in the majority of case by all three loss functions, while the AG tests leads to the equivalence between the two models more than 95% of the times. Overall, DGP1 appears to be a case where model misspecification is detectable both via the AG and DM tests.

Next consider DGP2, where the dynamic parameters are the same in the two groups. The loss functions L_1 and L_2 still prefer $M1$ to $M3$ more than 90% of the cases. However, in the comparison between $M1$ to $M4$ where both forms of misspecifications are included, the correctly specified model is selected only with frequency between 20% to 30% by all three loss function, and the misspecified model $M4$ is signaled as the preferred model by L_1 and L_2 in over 50% of the replications. The AG test has a similar behavior. The results for DGP2 show that there are configurations where misspecification is hard to detect via the AG and DM tests.

	sc	sc-M	sc-CS	dg	dg-M	dg-CS	res	full
C_{11}	0.09	0.11	0.10	0.26	0.41	0.37	*0.13*	*0.17*
C_{21}	0.03	0.06	0.05	0.11	*0.05*	0.07	*0.05*	*0.08*
C_{22}	0.10	0.12	0.11	0.30	0.17	0.21	0.24	*0.26*
C_{31}	0.03	0.05	0.04	0.06	*0.05*	0.06	*0.07*	*0.06*
C_{32}	0.03	0.06	0.05	0.06	0.06	0.10	0.13	*0.15*
C_{33}	0.08	0.10	0.09	0.09	0.09	0.11	*0.17*	*0.16*
C_{41}	0.02	0.04	0.03	*0.05*	*0.02*	*0.01*	*0.03*	*0.04*
C_{42}	0.03	0.05	0.03	0.05	*0.05*	0.04	*0.09*	*0.13*
C_{43}	0.04	0.06	0.06	0.04	0.05	0.06	0.09	*0.10*
C_{44}	0.07	0.08	0.06	0.05	0.05	0.04	*0.06*	*0.08*
C_{51}	0.04	0.06	0.05	0.08	*0.06*	0.02	*0.10*	*0.14*
C_{52}	0.03	0.06	0.04	0.09	0.06	0.07	*0.09*	*0.11*
C_{53}	0.02	0.04	0.03	0.03	0.03	0.04	0.07	*0.07*
C_{54}	0.01	0.03	0.02	*0.02*	0.02	*0.01*	*0.02*	*0.05*
C_{55}	0.08	0.10	0.09	0.14	0.13	0.12	*0.13*	0.15
C_{61}	0.02	0.05	0.02	*0.04*	*0.01*	0.04	*0.02*	*0.04*
C_{62}	0.02	0.05	0.03	0.05	*0.04*	0.05	0.05	*0.05*
C_{63}	0.02	0.04	*0.01*	*0.02*	*0.02*	*0.02*	*0.04*	*0.04*
C_{64}	*0.01*	0.03	*0.01*	*0.01*	*0.02*	*0.02*	*0.03*	*0.04*
C_{65}	0.02	0.04	0.02	*0.03*	*0.01*	*0.02*	*0.02*	*0.04*
C_{66}	0.07	0.09	0.07	0.05	0.05	0.05	*0.05*	0.05

TABLE 6. Elements of the symmetric matrix $\widehat{C}$; italics denote insignificant parameter estimates at the 0.01 level.

This limited Monte Carlo study shows that tests of equal predictive ability can help to select the appropriate weight matrices. The forecast criteria using loss functions L_i appear to favor parameter parsimony also when these restrictions are not present in the DGP, provided the true parameter values are not too heterogeneous.

7. An empirical application

This section presents an empirical application to six industrial stocks from the NYSE. The data, taken from Datastream, include logarithmic daily total returns of the following six Industrials companies: Honeywell International, Boeing, Caterpillar, General Dynamics, Raytheon, and Weyerhauser. We selected daily observations from January 2, 1997 to December 30, 2005 as estimation sample, and January to December 2006 as out-of-sample period. This gave 2265 observations in estimation and 251 observations out-of-sample. As a simple conditional mean specification, we fitted a constant to each series and considered the demeaned time series as u_t.

We estimated various BEKK specifications, see (2). The matrix C was always left unrestricted. For the parameter matrices A and B we considered various specifications, including structured ones. We considered the weight matrix $W^M = J_6 = \frac{1}{5}(1_6 1_6' - I_6)$, which represents a common sector-effect for all the six assets. We also defined the weight matrix W^C where neighbors correspond to companies active or competing in at least one subsector of the Industrial sector; for example,

	sc	sc-M	dg	dg-M		sc	sc-M	dg	dg-M
$\alpha_1^{(0)}$	0.200	0.200	0.309	0.458	$\beta_1^{(0)}$	0.974	0.973	0.932	0.853
$\alpha_2^{(0)}$			0.177	0.121	$\beta_2^{(0)}$			0.952	0.977
$\alpha_3^{(0)}$			0.231	0.234	$\beta_3^{(0)}$			0.966	0.967
$\alpha_4^{(0)}$			0.141	0.131	$\beta_4^{(0)}$			0.985	0.987
$\alpha_5^{(0)}$			0.328	0.314	$\beta_5^{(0)}$			0.942	0.945
$\alpha_6^{(0)}$			0.205	0.198	$\beta_6^{(0)}$			0.975	0.976
$\alpha_1^{(1)}$		0.024		-0.269	$\beta_1^{(1)}$		-0.012		0.108
$\alpha_2^{(1)}$				0.068	$\beta_2^{(1)}$				-0.019
$\alpha_3^{(1)}$				-0.060	$\beta_3^{(1)}$				*0.003*
$\alpha_4^{(1)}$				-0.038	$\beta_4^{(1)}$				*-0.001*
$\alpha_5^{(1)}$				-0.019	$\beta_5^{(1)}$				0.009
$\alpha_6^{(1)}$				*0.009*	$\beta_6^{(1)}$				*0.003*

TABLE 7. Parameter estimates of matrices $\widehat{A}$ and $\widehat{B}$ for selected specifications. Italics denote insignificant parameter estimates at the 0.01 level.

Honeywell and Boeing are both active in the Defence subsector. We also considered a weight matrix W^S which defines companies within the same supply chain as neighbors (e.g. companies characterised by vertical or horizontal integration). For example, Boeing and Caterpillar have joint contracts in the Defence subsector. The resulting matrices W^C and W^S were given by

$$W^C = \begin{pmatrix} & 1/3 & & 1/3 & 1/3 & \\ 1/3 & & & 1/3 & 1/3 & \\ & & & & & 1 \\ 1/2 & 1/2 & & & & \\ 1/2 & 1/2 & & & & \\ & & 1 & & & \end{pmatrix}, \qquad W^S = \begin{pmatrix} & & & 1 & & \\ & & & 1 & & \\ 1/3 & 1/3 & & & 1/3 & \\ & & & 1/2 & & 1/2 \\ & & & & 1 & \\ & & & & & \end{pmatrix},$$

where 0 entries are omitted for readability.

Using the weight matrices W^M, W^C, W^S, the following BEKK specifications were estimated:

- sc (scalar): $A = \alpha_1^{(0)} I_6$, $B = \beta_1^{(0)} I_6$;
- sc-M (scalar, structured with weight matrix W^M): $A = \alpha_1^{(0)} I_6 + \alpha_1^{(1)} W^M$, $B = \beta_1^{(0)} I_6 + \beta_1^{(1)} W^M$;
- sc-CS (scalar. structured with weight matrices W^C and W^S): $A = \alpha_1^{(0)} I_6 + \alpha_1^{(1)} W^C + \alpha_1^{(2)} W^S$, $B = \beta_1^{(0)} I_6 + \beta_1^{(1)} W^C + \beta_1^{(2)} W^S$;
- dg (diagonal): $A = \mathrm{diag}(\alpha^{(0)})$, $B = \mathrm{diag}(\beta^{(0)})$;
- dg-M (diagonal, structured with weight matrix W^M): $A = \mathrm{diag}(\alpha^{(0)}) + \mathrm{diag}(\alpha^{(1)}) W^M$, $B = \mathrm{diag}(\beta^{(0)}) + \mathrm{diag}(\beta^{(1)}) W^M$;
- dg-CS (diagonal, structured with weight matrices W^C and W^S): $A = \mathrm{diag}(\alpha^{(0)}) + \mathrm{diag}(\alpha^{(1)}) W^C + \mathrm{diag}(\alpha^{(2)}) W^S$, $B = \mathrm{diag}(\beta^{(0)}) + \mathrm{diag}(\beta^{(1)}) W^C + \mathrm{diag}(\beta^{(2)}) W^S$;
- res (restricted): the parameter matrices A and B have off-diagonal zero restrictions corresponding to 0 entries in $W^S + W^C$;
- full : A, B unrestricted.

The idea in the res specification is to impose no spillover and feedback effects across companies which are not economically linked. The parameters of the various specifications are (subsets) of

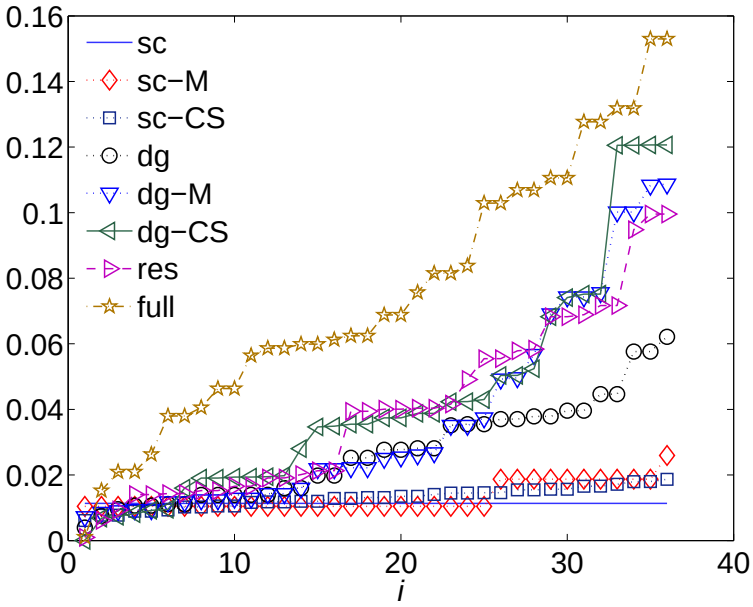

FIGURE 1. Ordered values of $|1-\lambda_i|$ where λ_i are the eigenvalues of $\widehat{A}\otimes\widehat{A}+\widehat{B}\otimes\widehat{B}$, $i=1,\ldots,36$.

the vectors $\alpha^{(j)}=(\alpha_1^{(j)}:\cdots:\alpha_6^{(j)})'$, $\beta^{(j)}=(\beta_1^{(j)}:\cdots:\beta_6^{(j)})'\in\mathbb{R}^6$, $j=0$, 1, and 2. Tables 6 and 7 report a subset of the estimated parameters, where italics indicate insignificant coefficients at the 1% level. Standard errors were calculated using the robust sandwich variance estimator, see Section 5.3.

Table 6 reports the elements in matrix $\widehat{C}$; it shows that all off-diagonal elements are significant for the scalar specifications; some coefficients are insignificant for the diagonal ones. In particular the dg-M model has many insignificant intercepts for the covariance terms, which could be further restricted to 0. This suggests that failing to account for spillover and feedback effects may induce spuriously significant intercept terms in some covariance dynamic equations. The full model and the res model, on the other hand, have many insignificant coefficient also on the diagonal of C; this may reflect the fact that these specifications are possibly over-parameterized.

Table 7 contains the coefficients in A, B for some structured specifications; it shows the changes induced by the introduction of the various weight matrices. Structured specifications estimates vary across assets, and they are generally significant. Estimated $\beta_i^{(j)}$coefficients are usually lower and the $\alpha_i^{(j)}$ coefficients are usually higher for the structured specifications.

The persistence in the conditional covariance dynamics is associated with the eigenvalues λ_i of $\widehat{A}\otimes\widehat{A}+\widehat{B}\otimes\widehat{B}$. In Fig. 1 we report the 36 ordered values of $|1-\lambda_i|$ where $|\cdot|$ indicates the modulus. It is seen that structured specifications are characterized by less dynamic persistence when compared with the corresponding unstructured restricted specifications. The full model shows the least persistence, even though the largest eigenvalues are closer to 1. This suggests that the more restricted models imply persistent dynamics in more dimension, a possibly undesirable feature.

Table 8 reports results on model fit in term of the (pseudo) Gaussian log-likelihood. The table reports the number of model parameters as well as the number of significant parameters using a 1% level. We observe that in larger models most parameters are not significant, while in more restricted models most coefficients are significant.

	log-Lik	par	sig. par.	res	dg-CS	dg-M	dg	sc-CS	sc-M	sc
full	-15128.3	93	14	-0.870	0.403	0.716	1.244	1.596	2.051	1.796
res	-15258.4	71	13		*2.915*	*3.267*	*4.508*	*4.626*	*5.358*	*4.925*
dg-CS	-15372.8	55	18			0.896	*2.764*	*3.097*	*4.137*	*3.596*
dg-M	-15396.9	45	20				1.627	2.452	*3.722*	*3.051*
dg	-15481.9	33	26					1.936	*3.940*	*3.052*
sc-CS	-15549.1	27	22						*7.369*	*4.092*
sc-M	-15556.0	25	25							*-5.013*
sc	-15564.1	23	20							

TABLE 8. AG test statistic computed with sandwich variance estimator. Italics identify cases where the null hypothesis is rejected at the 1% level. A negative sign denotes preference for the column model whereas a positive sign a preference for the row model. The first column reports the log-likelihood value, the second contains the number of parameters in the models, and the third column the number of statistically significant parameters at 1% level.

	res	dg-CS	dg-M	dg	sc-CS	sc-M	sc
full	0.661	*3.851*	*4.149*	*3.699*	*3.575*	*4.071*	*3.764*
res		*4.426*	*4.607*	*3.844*	*2.938*	*3.379*	*3.090*
dg-CS			0.816	0.481	0.823	1.345	1.028
dg-M				-0.295	0.504	1.119	0.744
dg					0.926	1.774	1.24
sc-CS						*6.964*	1.529
sc-M							*-3.215*

TABLE 9. DM test statistic computed with loss function L_1. Italics identify cases where the null hypothesis of equal performance is rejected at the 1% level. A negative sign denotes preference for the column model whereas a positive sign a preference for the row model.

We next compared models via the AG and DM forecast tests, as in Section 6. Because the true covariance matrix is not available, we replaced it with the outer product of realized returns in the calculations of the DM tests. This choice allows to compute L_1 and L_3 but not L_2, which requires Σ_t to be nonsingular. In the following we report results only for L_1; results for L_3 were similar.

Tables 8 and 9 report results for the AG and DM tests, where models are listed according to their number of parameters, from largest to smallest. For both the AG and DM tests, the larger full and res models are preferred. In models with a larger cross-section, it may not be possible to estimate them, due to the curse of dimensionality problem; in this case the structured specifications may be the largest estimable models. This situation corresponds to starting from the row corresponding to the dg-CS model; in this case, one would select the dg-CS structured specification as the preferred model.

To further compare the models' forecasting abilities, we also considered a portfolio allocation framework, along the lines of indirect model comparison methods, see Patton and Sheppard (2009), and Caporin and McAleer (2012). We consider two specific portfolios, the equally weighted (EW)

	res	dg-CS	dg-M	dg	sc-CS	sc-M	sc
full	1.477	*4.518*	*4.140*	*2.605*	2.304	*3.367*	*2.561*
res		*5.557*	*4.979*	*2.944*	2.221	*4.135*	2.500
dg-CS			-1.679	*-2.608*	-2.480	-1.439	-2.258
dg-M				-2.239	-2.272	-0.425	-2.028
dg					-1.337	*4.573*	-0.328
sc-CS						*4.778*	1.434
sc-M							*-3.542*

TABLE 10. AG test results for EW strategies. Italics identify cases where the null hypothesis of equal performance is rejected at the 1% level. A negative sign denotes preference for the column model whereas a positive sign a preference for the row model.

	res	dg-CS	dg-M	dg	sc-CS	sc-M	sc
full	-0.759	-0.458	-1.287	-2.106	-1.663	-0.508	-1.948
res		0.338	-0.928	*-2.789*	-1.974	0.502	-2.419
dg-CS			-0.936	-3.474	-1.712	0.086	-2.184
dg-M				-1.788	-0.781	1.201	-1.296
dg					1.632	*3.453*	0.935
sc-CS						*5.719*	*-2.687*
sc-M							*-6.272*

TABLE 11. AG test results for GMV strategies. Italics identify cases where the null hypothesis of equal performance is rejected at the 1% level. A negative sign denotes preference for the column model whereas a positive sign a preference for the row model.

portfolio (weights are equal to 1/6 for each asset), and the global minimum variance portfolio (GMV), with weights equal to $\Sigma_t^{-1}1_6(1_6'\Sigma_t^{-1}1_6)^{-1}$. When comparing the EW and the GMV portfolios, differences across models depend only on the dissimilarities across covariance specifications. For both the GMV and EW cases and for all specifications we computed the portfolio predicted variance as a measure of portfolio risk, and then we used this quantity to evaluate the AG test.

Results are collected in Tables 10 and 11. The results for EW and GMV strategies appear to be different. For EW strategies, the full and res model outperform the remaining models. In case these two larger models are not estimable, the medium size models are worse than the parsimonious structured sc-CS specification. This suggests that there is a prize for shrinkage in this case. Note that also here structured specifications are selected as the preferred model. For GMV strategies, the large models - full and res - are inferior to the more parsimonious specifications. In this case the selected specification is the sc-M model, again a structured specification.

An alternative approach to compare forecasts is to use a utility function. We assume a power utility function $U = (W^{1-\phi})/(1-\phi)$ with risk aversion parameter ϕ. We fix the initial wealth at $W = 1$ and graphically compare utilities over time, setting the wealth equal to the cumulated returns of the GMV portfolio, using each model prediction for Σ_t. The portfolios are rotated daily according to the covariance forecasts of each model, and we assume no transaction costs.

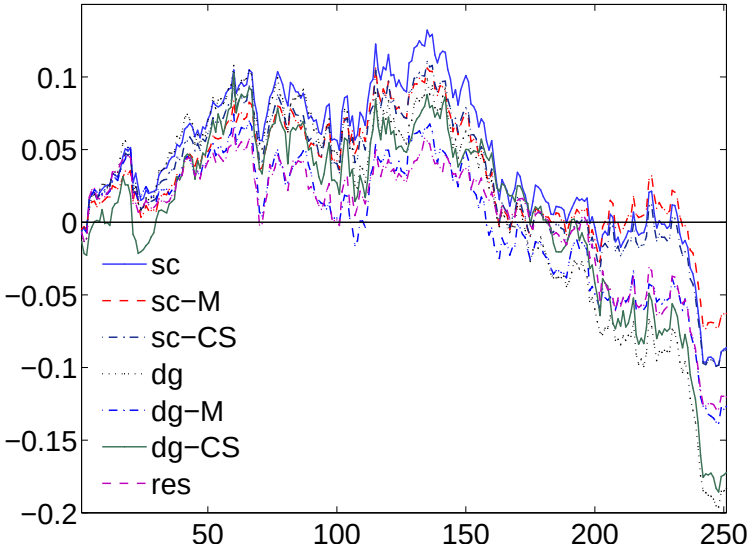

FIGURE 2. $\log(U_j/U_{\text{full}})$, where U_j is the utility associated with the cumulated wealth produced by GMV portfolio using the predictions of model j. The scale of the graph is for CRRA parameter equal to 5; different CRRA values simply imply a re-scaling and do not affect the relative performance of the models.

Figure 2 reports the log of the ratio between the utilities of each model and the utility associated with the predictions of the full model.[12] We note the absence of a stable ordering across models. This evidence is in line with the Monte Carlo evidence in Clements, Doolan, Hurn, and Becker (2009), who report that economic loss functions that rely on portfolio returns and on investor utility have weaker power to distinguish between competing forecasts. Overall, the empirical results suggest that the proposed structured specifications are a promising modelling option, especially for situations when large models are not estimable due to the curse of dimensionality problem.

8. Conclusions

In this paper we have shown how structured specification can be defined in a number of MVM models, using weight matrices to condense information coming from other returns series. Structured specifications form an interesting modelling option for volatility models, because they provide both flexible and parsimonious parameterizations, allowing for variance spillover and feedback effects. Moreover, they are characterized by a number of parameters that grows linearly with the cross-sectional dimension; at the same time, parameters have a direct economic interpretation that reflects the chosen notion of economic proximity. An empirical application shows how the relative performance of the different specification depends on the criterion used in the forecast evaluation and it highlights the potential of the structured specification, especially for situations when large models are not estimable, due to the curse of dimensionality problem.

Acknowledgement

Partial financial support from Italian MUR Grant Cofin2006-13-1140 (both authors) and Danish Social Science Research Council Grant 2114-04-0001 (second author) are gratefully acknowledged.

[12]The utility function depends on a risk-aversion coefficient, which only impacts the scale of the graph, but not the ordering of the models. In fact, $\log U = (1-\phi)\log W - \log(1-\phi)$, and hence $\log(U_j/U_{\text{full}}) = (1-\phi)\log(W_j/W_{\text{full}})$.

We thank, without implicating, the following people for useful comments on previous versions of the paper: the Editor Esfandiar Maasoumi, two anonymous referees, John Aldrich, Christian Brownlees, Niels Haldrup, Grant Hillier, Joel Hasbrouck, Søren Johansen, Federico Martellosio, Michael McAleer, Rasmus Søndergaard Pedersen, Christophe Planas, Anders Rahbek, Timo Teräsvirta, as well as several participants to paper presentations.

References

Abadir, K., W. Distaso, and F. Zikes (2010), Model-free estimation of large covariance matrices. Discussion paper, 10-17, The Rimini Centre for Economic Analysis, available at www.rcfea.org/RePEc/pdf/wp17_10.pdf.

Amisano, G., and R. Giacomini (2007), Comparing Density Forecasts via Weighted Likelihood Ratio Tests. Journal of Business & Economic Statistics, 25, 177–190.

Anselin, L. (1988), Spatial Econometrics: Methods and Models. Kluwer Academic Publishers.

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

Baltagi, B. H., G. Bresson, and A. Pirotte (2007), Panel unit root tests and spatial dependence. Journal of Applied Econometrics, 22(2), 339–360.

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

Billingsley, P. (1961), Statistical Inference for markow Processes. University of Chicago Press.

Bollerslev, T., and J. Wooldridge (1992), Quasi-Maximum Likelihood Estimation and Inference in Dynamic Models with Time Varying Covariances. Econometric Reviews, 11, 143–172.

Burridge, P., and B. Fingleton (2010), Bootstrap Inference in Spatial Econometrics: the J-test. Spatial Economic Analysis, 5, 93–119.

Caporin, M., and M. McAleer (2012), Robust Ranking of Multivariate GARCH Models by Problem Dimension. Forthcoming in Computational Statistics and Data Analysis, available as KIER Working Papers 815, Kyoto University, Institute of Economic Research.

Caporin, M., and P. Paruolo (2005a), Multivariate ARCH with spatial effects for stocks sector and size. WP 2005/13, available at http://eco.uninsubria.it/dipeco/qf/2005.htm, UniversitÃ dell'Insubria, Varese, Italy.

——— (2005b), Spatial effects in multivariate ARCH. WP 2005/1, available at http://eco.uninsubria.it/dipeco/qf/2005.htm, UniversitÃ dell'Insubria, Varese, Italy.

Case, A. (1991), Spatial patterns in household demand. Econometrica, 59, 953–966.

Clements, A., M. Doolan, S. Hurn, and M. Becker (2009), Evaluating multivariate volatility forecasts. Discussion paper, NCER Working Paper Series, number 41.

Comte, F., and O. Lieberman (2003), Asymptotic theory for multivariate GARCH processes. Journal of Multivariate Analysis, 84, 61–84.

Conley, T. G., and B. Dupor (2003), A Spatial Analysis of Sectoral Complementarity. Journal of Political Economy, 111.

Cressie, N. (1993), Statistics for spatial data, revised edition,. Wiley, New York.

Davidson, R., and J. G. MacKinnon (1993), Estimation and Inference in Econometrics. Oxford University Press.

Diebold, F. X., and R. S. Mariano (1995), Comparing Predictive Accuracy. Journal of Business & Economic Statistics, 13(3), 253–63.

Engle, R., and B. Kelly (2012), Dynamic Equicorrelations. Journal of Business & Economic Statistics, 30, 212–228.

Engle, R. F., and F. K. Kroner (1995), Multivarariate simultaneous generalized ARCH. Econometric Theory, 11, 122–150.

Francq, C., and J. M. Zakoian (2010), QML estimation of a class of multivariate GARCH models without moment conditions on the observed process. Discussion paper, MPRA paper 20779 University Library of Munich, Germany, available at `http://mpra.ub.uni-muenchen.de/20779/`.

Gourieroux, C., J. Jasiak, and R. Sufana (2009), The Wishart autoregressive process of multivariate stochastic volatility. Journal of Econometrics, 150, 167–181.

Hafner, C. M., and A. Preminger (2009a), Asymptotic theory for a factor GARCH model. Econometric theory, 25, 336–363.

——— (2009b), On asymptotic theory for multivariate GARCH models. Journal of Multivariate Analysis, 100, 2044–2054.

James, W., and C. Stein (1961), Proceedings of the Fourth Berkley Symposium on Mathematics Statistics and Probability, Chapter: Estimation with Quadratic Loss, pp. 361–379. University of California Press, Berkeley, California.

Kelejian, H. H. (2008), A spatial J-test for model specification against a single or a set of non-nested alternatives. Letters in Spatial and Resource Sciences, 1, 3–11.

Laurent, S., J. Rombouts, and F. Violante (2012), On the forecasting accuracy of multivariate GARCH models. Journal of Applied Econometrics, 27, 934–955.

Ling, S., and M. McAleer (2003), Asymptotic theory for a vector ARMA-GARCH model. Econometric Theory, 18, 280–310.

Linton, O. (2009), Semiparametric and Nonparametric ARCH Modelling. In Handbook of Financial Time Series, ed. by T. Andersen, R. Davis, J. Kreiss, and T. Mikosch, pp. 157–168. Springer.

Lucchetti, R. (2006), Identification of covariance structures. Econometric Theory, 22, 235–257.

Magnus, J. (1988), Linear Structures. Oxford University Press.

Magnus, J., and H. Neudecker (2007), Matrix Differential Calculus with Applications in Statistics and Econometrics, 3rd ed. Wiley.

Magnus, J. R. (1982), Multivariate error component analysis of linear and nonlinear regression models by maximum likelihood. Journal of Econometrics, 19, 239–285.

Martellosio, F. (2010), Power properties of invariant tests for spatial autocorrelation in linear regression. Econometric Theory, 26, 152–186.

McAleer, M. (2005), Automated inference and learning in modeling financial volatility. Econometric Theory, 21, 232–261.

McAleer, M., F. Chan, S. Hoti, and O. Lieberman (2008), Generalized autoregressive conditional correlation. Econometric Theory, 24, 1554–1583.

McAleer, M., and M. Medeiros (2008), A multiple regime smooth transition Heterogeneous Autoregressive model for long memory and asymmetries. Journal of Econometrics, 147, 104–119.

McAleer, M., M. Medeiros, and D. Slottje (2008), A neural network demand system with heteroskedastic errors. Journal of Econometrics, 147(2), 359 – 371.

Paruolo, P. (1997), Asymptotic inference on the moving average impact matrix in cointegratared I(1) VAR systems. Econometric Theory, 13, 79–118.

——— (1998), Tests of integration in circular autoregressive models. Statistical Methods and Applications - Journal of the Italian Statistical Society, 7, 297–306.

Patton, A., and K. Sheppard (2009), Evaluating Volatility and Correlation Forecasts. In Handbook of Financial Time Series, ed. by T. Andersen, R. Davis, J. Kreiss, and T. Mikosch, pp. 801–838. Springer.

Pesaran, M. H., T. Schuermann, and S. Weiner (2004), Modelling regional interdependencies using a global error-correcting macroeconometric model. Journal of Business Economics and Statistics, 22, 129–162.

Pesaran, M. H., and E. Tosetti (2011), Large panels with common factors and spatial correlation. Journal of Econometrics, 161(2), 182 – 202.

Rothenberg, T. J. (1971), Identification in Parametric Models. Econometrica, 39, pp. 577–591.

Seneta, E. (1981), Non-Negative Matrices and Markov Chains. Springer.

Silvapulle, M. J., and P. Silvapulle (1995), A score test against one-sided alternatives. Journal of the American Statistical Association, 90, 342–349.

Silvennoinen, A., and T. Teräsvirta (2009), Multivariate GARCH models. In Handbook of Financial Times Series, ed. by T. Andersen, R. Davis, J. Kreiss, and T. Mikosch, pp. 201–229. Springer.

Wichura, M. J. (2006), The coordinate-free approach to linear models. Cambridge University Press, Cambridge.

A. Appendix: proximity matrices

This Appendix provides definitions and properties of proximity matrices, as well as the proof of Theorem 1 in Section 4. In the following Definition 7 of proximity matrices $\mathcal{S}_n$ weight matrices can be taken to be row-normalized without loss of generality; we hence assume that weight matrices are row-normalized. We employ the following notation: $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_{0+}$ are the set of all, positive and nonnegative real numbers; $\mathbb{J} := \{0, 1\}$ is the binary set containing only 0 and 1; $\mathbb{A}^{n\times m}$ indicates the set of all matrices of dimensions $n \times m$ whose entries belong to the set $\mathbb{A}$; $\mathbb{A}^n := \mathbb{A}^{n\times 1}$.

Definition 6 (Classes of weight and stochastic matrices)**.** *Define the following classes of $n \times n$ matrices:*

- *the class of extended stochastic matrices* $\mathcal{U}_n := \{W \in \mathbb{R}_{0+}^{n\times n} : W1_n \in \mathbb{J}^n, W \neq 0\}$;
- *the class of extended weight matrices* $\mathcal{V}_n := \{W \in \mathcal{U}_n : \mathrm{dg}(W) = 0\}$;
- *the class of stochastic matrices* $\mathcal{P}_n := \{W \in \mathcal{U}_n : W1_n = 1_n\}$;
- *the class of weight matrices* $\mathcal{W}_n := \mathcal{P}_n \cap \mathcal{V}_n$.

Observe that Definition 6 implies $\mathcal{V}_n, \mathcal{P}_n \subset \mathcal{U}_n$ and $\mathcal{W}_n \subset \mathcal{P}_n, \mathcal{V}_n$. We next define the class of proximity matrices $\mathcal{S}_n$ as (a generalization of) the class containing linear combinations of the identity and weight matrices $\mathbb{W} := \{W_i\}_{i=1}^k$ with $W_i \in \mathcal{W}_n$ or $\mathcal{V}_n$.

Definition 7 (Proximity matrices)**.** *Let $\mathbb{W}:=\{W_i\}_{i=0}^k$ with $W_0 := I_n$ and $W_i \in \mathcal{W}_n$ or $\mathcal{V}_n$ be a given set of (possibly extended) weight matrices for $i \geq 1$. We define the class of proximity matrices $\mathcal{S}(\mathbb{W})$ defined as the set $\{A \in \mathbb{R}^{n\times n} : A = \sum_{i=0}^k A_i W_i, A_i = \mathrm{diag}(a_i), a_i \in \mathbb{R}^n, \forall i\}$. We indicate the subclass of homogeneous proximity matrices $\mathcal{S}_n^H(\mathbb{W})$ as the elements in $\mathcal{S}_n(\mathbb{W})$ with $a_i = \alpha_i 1_n$ for all i.*

In the following we omit the argument $\mathbb{W}$ in $\mathcal{S}_n$ unless needed for clarity. Note that by construction $\mathcal{S}^H \subset \mathcal{S}$ and that if $A \in \mathcal{A}_n$ then $I - A \in \mathcal{A}_n$, $\mathcal{A}_n = \mathcal{S}_n^H, \mathcal{S}$. We next consider convex combinations and product of weight matrices. The latter is motivated by the fact that in some special cases, given a certain weight matrix W that represents first-order neighbors, W^j represents j-th order neighbors, see e.g. Paruolo (1998). The former is of interest because one may wish to consider an 'average' weight matrix $c_1W_1 + c_2W_2$ of two weight matrices W_1 and W_2, with scalar coefficients c_j. It turns out, see the following Theorem 8, that W^j are not in general weight matrices while $c_1W_1 + c_2W_2$ is.

Formally, we say that a class $\mathcal{A}$ is closed with respect to convex combinations if $\sum_{i=1}^k c_iW_i \in \mathcal{A}$ for $0 \leq c_i \leq 1$, $\sum_{i=1}^k c_i = 1$ when $W_i \in \mathcal{A}$ for all $i = 1, \ldots, k$. Similarly we say that a class $\mathcal{A}$ is closed with respect to matrix multiplication if $\prod_{i=1}^k W_i \in \mathcal{A}$ when $W_i \in \mathcal{A}$ for all $i = 1, \ldots, k$.

Theorem 8 (Convex combinations and matrix products)**.** *Let $\mathcal{A}_n$ indicate one of the classes $\mathcal{U}_n$, $\mathcal{V}_n$, $\mathcal{P}_n$, $\mathcal{W}_n$, and let $W_i \in \mathcal{A}_n$, $i = 1, \ldots, k$. Then the following properties hold:*
(i) $\mathcal{A}_n$ is closed with respect to convex combinations for $\mathcal{A}_n = \mathcal{W}_n$, $\mathcal{P}_n$ but not for $\mathcal{A}_n = \mathcal{U}_n$, $\mathcal{V}_n$.
(ii) $\mathcal{A}_n$ is closed with respect of matrix multiplication for $\mathcal{A}_n = \mathcal{P}_n$ but not for $\mathcal{A}_n = \mathcal{U}_n$, $\mathcal{V}_n$, $\mathcal{W}_n$.

Proof of Theorem 8. Both statements are well known for $\mathcal{A}_n = \mathcal{P}_n$, see e.g. Seneta (1981), Chapter 4; hence we only consider the remaining classes.
Proof of (i). Let $\mathcal{A}_n = \mathcal{W}_n$; one has $W := \sum_{i=1}^k c_iW_i \in \mathbb{R}_{0+}^{n\times n}$. Post-multiplying by 1_n one has $\sum_{i=1}^k c_iW_i1_n = \sum_{i=1}^k c_i1_n = (\sum_{i=1}^k c_i)1_n = 1_n$. Similarly one finds $\mathrm{dg}(W) = \sum_{i=1}^k c_i \mathrm{dg}(W_i) = \sum_{i=1}^k c_i0 = 0$, which completes the proof for $\mathcal{A}_n = \mathcal{W}_n$.
A counterexample for $\mathcal{A}_n = \mathcal{V}_n$ is given by $W_1 = (0 : e_1)$, $W_2 = (e_2 : 0)$, $c_1 = c_2 = \frac{1}{2}$, for which $W := \sum_{i=1}^2 c_iW_i = \frac{1}{2}(e_2 : e_1)$; despite W_1, $W_2 \in \mathcal{V}_2$, $W1_2 = \frac{1}{2}1_2 \notin \mathbb{J}^2$, and hence $W \notin \mathcal{V}_2$. This also implies that (i) does not hold for $\mathcal{A}_n = \mathcal{U}_n$, given that $\mathcal{V}_n \subset \mathcal{U}_n$.
Proof of (ii). A counterexample for $\mathcal{A}_n = \mathcal{W}_n$ is obtained choosing $W_1 = W_2 = J_3 = \frac{1}{2}(1_31_3' - I_3)$, for which $W := W_1W_2 = \frac{1}{4}(1_31_3' + I_3)$, so that $\mathrm{dg}(W) = \frac{1}{2}1_3 \neq 0$ and hence $W \notin \mathcal{W}_n$ despite W_1,$W_2 \in \mathcal{W}_n$. This also implies that (ii) does not hold for $\mathcal{A}_n = \mathcal{V}_n$ because $\mathcal{W}_n \subset \mathcal{V}_n$.
A counterexample for $\mathcal{A}_n = \mathcal{U}_n$ is given by $W_1 = W_2 = \frac{1}{2}(e_2 : e_1 : e_1 + e_2) \in \mathcal{U}_n$; one finds $W := W_1W_2 = \frac{1}{4}(e_1 : e_2 : e_1 + e_2)$ so that $W1_3 = \frac{1}{2}(e_1 + e_2) \notin \mathbb{J}^3$ and hence $W \notin \mathcal{U}_3$. □

We next analyze properties of Kronecker products of weight matrices in the following Theorem 9. A motivation for this interest is given by the observation that when x_{jt}, $j = 1, 2$, are independent Markov Chains (MC) with n discrete states and transition probabilities given by the stochastic matrices $P_j \in \mathcal{P}_n$, then (x_{1t}, x_{2t}) is still a MC with n^2 discrete states and transition probabilities given by the entries in $P_1 \otimes P_2$, when the ordered pairs (l, m) are placed in lexicographic order. This observation shows that $P_1 \otimes P_2 \in \mathcal{P}_{n^2}$ and the next theorem states, inter alia, that this is true also substituting $\mathcal{P}_n$ with $\mathcal{W}_n$.

Theorem 9 (Kronecker products of weight matrices)**.** *Let $\mathcal{A}_n$, $\mathcal{B}_n$ indicate one of the classes $\mathcal{U}_n$, $\mathcal{V}_n$, $\mathcal{P}_n$, $\mathcal{W}_n$; moreover let H, K denote generic $n \times n$ matrices; then the following properties of Kronecker products hold.*
(i) Let $W_1 \in \mathcal{A}_n$, $W_2 \in \mathcal{B}_n$; then $W_1 \otimes W_2 \in \mathcal{B}_{n^2}$ and $W_2 \otimes W_1 \in \mathcal{B}_{n^2}$ for $(\mathcal{A}_m, \mathcal{B}_m) = (\mathcal{U}_m, \mathcal{U}_m)$, $(\mathcal{U}_m, \mathcal{V}_m)$, $(\mathcal{V}_m, \mathcal{V}_m)$, $(\mathcal{P}_m, \mathcal{P}_m)$, $(\mathcal{P}_m, \mathcal{W}_n)$, $(\mathcal{W}_m, \mathcal{W}_m)$.
(ii) Conversely, let $H \otimes K \in \mathcal{B}_{n^2}$ or $K \otimes H \in \mathcal{B}_{n^2}$; this implies $H = cW_1$ and $K = \frac{1}{c}W_2$, $c \in \mathbb{R}\backslash\{0\}$, where $W_1 \in \mathcal{A}_n$,$W_2 \in \mathcal{B}_n$ for $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{U}_n, \mathcal{U}_n)$, $(\mathcal{U}_n, \mathcal{V}_n)$, $(\mathcal{P}_n, \mathcal{P}_n)$, $(\mathcal{P}_n, \mathcal{W}_n)$.

Proof of Theorem 9(*i*). Consider $W := W_1 \otimes W_2$, and observe that $W \in \mathbb{R}_{0+}^{n^2 \times n^2}$ for any pair $(\mathcal{A}_n, \mathcal{B}_n)$ in the statement. Let $a_i := W_i 1_n$, $i = 1, 2$; one finds

$$W 1_{n^2} = (W_1 \otimes W_2) 1_{n^2} = \operatorname{vec}(W_2 1_n 1_n' W_1') = \operatorname{vec}(a_2 a_1') =: b. \tag{25}$$

If $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{P}_n, \mathcal{P}_n)$, $(\mathcal{P}_n, \mathcal{W}_n)$ then $a_i = 1_n$ $i = 1, 2$ and hence $b = \operatorname{vec} 1_n 1_n' = 1_{n^2}$. This proves (i) for $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{P}_n, \mathcal{P}_n)$. When $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{P}_n, \mathcal{W}_n)$ we also need to check that $\operatorname{dg} W = 0$; this follows because

$$\operatorname{dg}(W_1 \otimes W_2) = \operatorname{dg} W_1 \otimes \operatorname{dg} W_2 \tag{26}$$

and $\operatorname{dg} W_2 = 0$ because $W_2 \in \mathcal{W}_n$. The same argument holds when $W_1 \in \mathcal{W}_n$ and $W_2 \in \mathcal{P}_n$, simply using $\operatorname{dg} W_1 = 0$. This proves (i) for $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{P}_n, \mathcal{W}_n)$ and also for $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{W}_n, \mathcal{W}_n)$.

Next consider (25) when $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{U}_n, \mathcal{U}_n)$, $(\mathcal{U}_n, \mathcal{V}_n)$, and hence $a_i \in \mathbb{J}^n$, $i = 1, 2$. This implies $a_2 a_1' \in \mathbb{J}^{n \times n}$ and $b \in \mathbb{J}^{n^2}$, which proves (i) for $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{U}_n, \mathcal{U}_n)$. When $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{U}_n, \mathcal{V}_n)$ we also need to check that $\operatorname{dg} W = 0$; this follows by (26) noting that $\operatorname{dg} W_2 = 0$ because $W_2 \in \mathcal{V}_n$. The same argument holds when $W_1 \in \mathcal{V}_n$ and $W_2 \in \mathcal{U}_n$, simply using $\operatorname{dg} W_1 = 0$. This proves (i) for $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{U}_n, \mathcal{V}_n)$, $(\mathcal{V}_n, \mathcal{V}_n)$.

Proof of (ii). Consider

$$(H \otimes K) 1_{n^2} = \operatorname{vec}(K 1_n 1_n' H) =: \operatorname{vec} A =: \operatorname{vec}(a_1 a_2') =: b, \tag{27}$$

where $a_1 := K 1_n$, $a_2 := H 1_n$. We note that $A := a_1 a_2'$ has rank 1 with rank decomposition $A = b_1 b_2'$ which satisfies $b_1 = c a_1$, $b_2 = c^{-1} a_2$ for a nonzero real scalar c.

If $H \otimes K \in \mathcal{B}_{n^2}$, for $\mathcal{B}_n = \mathcal{U}_n, \mathcal{V}_n$ then $b \in \mathbb{J}^{n^2}$ by definition. This implies that b_1 and b_2 can be chosen in $\mathbb{J}^n$, so that one has $b_1 := c^{-1} a_1$, $b_2 := c a_2 \in \mathbb{J}^n$. This shows that $c^{-1} K 1_n, c H 1_n \in \mathbb{J}^n$, and proves (ii) for $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{U}_n, \mathcal{U}_n)$. We also observe that $\operatorname{dg}(H \otimes K) = 0$ implies either $\operatorname{dg} H = 0$ or $\operatorname{dg} K = 0$ or both by (26); this proves (ii) for $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{U}_n, \mathcal{V}_n)$.

If $H \otimes K \in \mathcal{B}_{n^2}$ with $B_{n^2} = \mathcal{P}_{n^2}, \mathcal{W}_{n^2}$, then $b = 1_{n^2}$ and a_1 and a_2 can be chosen equal to $c^{-1} 1_n$, $c 1_n$. This shows that $c^{-1} K 1_n, c H 1_n = 1_n$. This proves (ii) for $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{P}_n, \mathcal{P}_n)$.

We also observe that $\operatorname{dg}(H \otimes K) = 0$ implies either $\operatorname{dg} H = 0$ or $\operatorname{dg} K = 0$ or both thanks to (26). This proves (ii) for $(\mathcal{A}_n, \mathcal{B}_n) = (\mathcal{P}_n, \mathcal{W}_n)$. Finally note that the same arguments used to prove (ii) can be applied interchanging the order of H and K in the Kronecker product. □

The next proposition states similar properties for $\mathcal{S}, \mathcal{S}^H$, which are the key in proving that a structured BEKK is a special case of a structured VEC specification.

Theorem 10 (Kronecker products of proximity matrices)**.** *Let $A, B \in \mathcal{A}_n(\mathbb{W})$, where $\mathcal{A}_n$ indicates one of the classes $\mathcal{S}$, $\mathcal{S}^H$ and $\mathbb{W} := \{W_i\}_{i=0}^k$ with $W_0 = I_n$ and $W_i \in \mathcal{W}_n$ or $\mathcal{V}_n$ for $i \geq 1$; then $A \otimes B \in \mathcal{A}_{n^2}(\mathbb{W}^*)$ with $\mathbb{W}^* := \{W_h^*\}_{h=0}^m$, $W_h^* := W_i \otimes W_j$, $h := (k+1)i + j$, $m := (k+1)^2 - 1$, where $W_0^* = I_{n^2}$ and $W_h^* \in \mathcal{W}_{n^2}$ or $\mathcal{V}_{n^2}$ for $h \geq 1$.*

Proof of Theorem 10. Let $A = \sum_{i=0}^k A_i W_i$ and $B = \sum_{i=0}^k B_i W_i$ be the representations of A and B in terms of the set of weight matrices $\mathbb{W} := \{W_i\}_{i=0}^k$. One has

$$A \otimes B = \sum_{i,j=0}^{k} A_i W_i \otimes B_j W_j = \sum_{i,j=0}^{k} (A_i \otimes B_j)(W_i \otimes W_j) = \sum_{h=0}^{m} C_h W_h^*, \tag{28}$$

where $C_h := A_i \otimes B_j$ and $W_h^* := W_i \otimes W_j$ for $h = (k+1)i + j$. By Theorem 9(*i*), one has $W_h^* \in \mathcal{W}_{n^2}$ or $\mathcal{V}_{n^2}$ unless $h = 0$, for which $W_0^* = W_0 \otimes W_0 = I_n \otimes I_n = I_{n^2}$. Hence $\mathbb{W}^* := \{W_h^*\}_{h=0}^m$ is a set of weight matrices in $\mathcal{W}_{n^2}$ or $\mathcal{V}_{n^2}$ for $h \geq 1$. This proves the statement when $A, B \in \mathcal{S}^H(\mathbb{W})$, i.e.

if $A_i = \alpha_i I_n$ and $B_i = \beta_i I_n$ are scalar matrices. In order to prove the statement for $A, B \in \mathcal{S}(\mathbb{W})$ let $A = \sum_{i=0}^k A_i W_i$ and $B = \sum_{i=0}^k B_i W_i$ with $A_i =: \operatorname{diag}(a_i)$, $B_i =: \operatorname{diag}(b_i)$, and denote $c_h := a_i \otimes b_j \in \mathbb{R}^{n^2}$. One has $C_h := A_i \otimes B_j = \operatorname{diag}(a_i) \otimes \operatorname{diag}(b_j) = \operatorname{diag}(a_i \otimes b_j) =: \operatorname{diag}(c_h)$, and hence $A \otimes B$ in (28) belongs to the class $\mathcal{S}(\mathbb{W}^*)$, see Definition 7. □

We end this subsection with the proof of Theorem 1, which specializes the result of Theorem 10 to the case of the matrices in Subsection 4.3.

Proof of Theorem 1 . Element w^*_{hv} in (11) is equal to $1(F_i = F_l, i \neq l)1(F_j = F_m, j \neq m)$ and

$$\begin{aligned}\sum_{v=1}^{n^2} w^*_{hv} &= \sum_{l=1}^{n}\sum_{m=1}^{n} 1(F_i = F_l, i \neq l)1(F_j = F_m, j \neq m) \\ &= \sum_{l=1}^{n} 1(F_i = F_l, i \neq l) \sum_{m=1}^{n} 1(F_j = F_m, j \neq m) = (n_i - 1)(n_j - 1).\end{aligned}$$

Hence

$$w_{hv} = \frac{w^*_{hv}}{\sum_{v=1}^{n^2} w^*_{hv}} = \frac{1}{(n_i - 1)} 1(F_i = F_l, i \neq l) \frac{1}{(n_j - 1)} 1(F_j = F_m, j \neq m).$$

On the other hand $(W_n \otimes W_n)_{hv} = w^{\diamondsuit}_{il} w^{\diamondsuit}_{jm}$ where $w^{\diamondsuit}_{ij} := (W)_{ij}$; hence, using (8), one sees that $w_{hv} = w^{\diamondsuit}_{il} w^{\diamondsuit}_{jm}$. This proves that $W^{(3)}_{n^2} = W_n \otimes W_n$. The proofs for $W^{(j)}_{n^2}$, $j = 1, 2$ follow along similar lines. Finally the proof of the statement concerning the general Kronecker product $H \otimes K$ is given in Theorem 10 above. □

B. Appendix: identification

This Appendix reports proofs of Theorems 3 and 5 of Section 5. *Proof of Theorem* 3. We indicate by Π_j, ψ_j and φ_j different values of the parameters Π, ψ and φ, $j = 1, 2$. Because Π is identified by assumption, one has $\Pi_1 - \Pi_2 \neq 0$ implies $\ell_1 \neq \ell_2$ with positive probability, where ℓ_j is the log-likelihood function evaluated at Π_j.

We next show that $\Pi_1 - \Pi_2 \neq 0$ if and only if $\psi_1 - \psi_2 \neq 0$ under the condition (14) that M has full column rank. In fact $\operatorname{vec} \Pi_j' = M\psi_j + c$, so that $\operatorname{vec}(\Pi_1 - \Pi_2)' = M(\psi_1 - \psi_2)$. Hence when M has full column rank $\Pi_1 - \Pi_2 \neq 0$ if and only if $\psi_1 - \psi_2 \neq 0$, which shows that ψ is identified. Conversely, if M does not have full column rank, one has that $0 = M(\psi_1 - \psi_2)$ for all $a := \psi_1 - \psi_2$ that lie in the right null space of M. Hence ψ_1 is observationally equivalent to all vectors of the form $\psi_1 + a$ with a in this null space, and hence ψ is not identified.

The proof for φ is identical to the one of ψ, replacing M with MH. □

Proof of Theorem 5. Observe that $C = S^{-1}VS^{-1\prime}$ can be written as

$$SCS' = DD', \tag{29}$$

where $DD' = V$, $D = \operatorname{diag}(\sigma)$, $v_i = \sigma_i^2 > 0$. This shows that the specification $C = S^{-1}VS^{-1\prime}$ is a special case of the so-called A-B model in structural VAR models, see Lucchetti (2006) and references therein. Observe also that, by taking the vec of S in (18) and of D, one has

$$\operatorname{vec}(S) = \mathcal{K}_n \operatorname{vec}(S') = \mathcal{K}_n M\psi + \operatorname{vec} I, \qquad \operatorname{vec} D = G\sigma. \tag{30}$$

One then applies the results in section 4.1 in Lucchetti (2006); the rank condition (21) here is simply a restatement of his eq. (26) on page 248. The order conditions are obtained requiring that the number of restricted parameters is not greater than the number of unrestricted parameters. □