UNIVERSITÀ DEGLI STUDI DI PADOVA

Dipartimento di Scienze Economiche "Marco Fanno"

STRUCTURED MULTIVARIATE VOLATILITY MODELS

MASSIMILIANO CAPORIN
University of Padova

PAOLO PARULO
University of Insubria

January 2009

"MARCO FANNO" WORKING PAPER N.91

Structured Multivariate Volatility Models*

Massimiliano Caporin
Università degli Studi di Padova, Italy
Dipartimento di Scienze Economiche
'Marco Fanno'

Paolo Paruolo
EC Joint Research Centre &
Università dell'Insubria, Varese Italy
Dipartimento di Economia

First draft: January 2008
Current version: December 2008

Abstract

This paper proposes structured parametrizations for multivariate volatility models, which use spatial weight matrices induced by economic proximity. These structured specifications aim at solving the curse of dimensionality problem, which limits feasibility of model-estimation to small cross-sections for unstructured models. Structured parametrizations possess the following four desirable properties: i) they are flexible, allowing for covariance spill-over; ii) they are parsimonious, being characterized by a number of parameters that grows only linearly with the cross-section dimension; iii) model parameters have a direct economic interpretation that reflects the chosen notion of economic classification; iv) model-estimation computations are faster than for unstructured specifications. We give examples of structured specifications for multivariate GARCH models as well as for Stochastic- and Realized-Volatility models. The paper also discusses how to construct spatial weight matrices that are time-varying and possibly derived from a set of covariates.

Keywords: MGARCH, Stochastic Volatility, Realized Volatility, Spatial models, ANOVA.

J.E.L. Classification: C31, C32, G11.

*Paper presented at the second ETSERN meeting in Copenhagen, and at CREATES in Århus, November 2008. Partial financial support from Italian MUR Grant Cofin2006-13-1140 is gratefully acknowledged. We thank the following people for useful comments on a previous version of the paper: Søren Johansen, Federico Martellosio, Michael McAleer, Christophe Planas, Timo Teräsvirta, as well as several participants to paper presentations.

Contents

1 Introduction

Multivariate volatility models (MVM) are used for asset pricing, portfolio selection, option pricing, hedging and risk management, see e.g. Bauwens et al. (2006). MVM include multivariate GARCH specifications (MGARCH, see the reviews in McAleer 2005, Bauwens et al. 2006, Silvennoinen and Teräsvirta 2008 and the glossary in Bollerslev 2008), Multivariate Stochastic Volatility models (MSV, see e.g. Asai, McAleer and Yu 2006), and Multivariate Realized Covariance models (MRV, see e.g. McAleer and Medeiros 2008).

A major challenge in MVM is the rapid increase of number of parameters as the cross-sectional dimension increases. Large cross-sections would be of interest in typical applications of MVM; this contrasts with the empirical practice, where these models are estimated in just a handful of dimensions or – only for very restrictive specifications – in medium-sized cross-sections.

In unrestricted MVM, 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 unfeasible. This problem is similar to the deterioration of the rate of convergence for nonparametric estimators in higher dimensions, see e.g. Linton (2008), and we refer to it as the 'curse of dimensionality' problem for MVM.[1]

In this paper we discuss 'structured' – i.e. restricted – MVM specifications which provide a possible solution to this problem. The structure we consider is formalized through spatial concepts, where proximity is induced by economic common factors, through the definition of weight matrices. Several approaches for the definition of weight matrices are given, which translate (possibly time-varying) past information into proximity coefficients.

The contributions of this paper consist in introducing spatial concepts and tools in the specifications of MVM. 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. This restricts the number of parameters, while also retaining a simple interpretation for coefficients.

Many empirical studies assume diagonal parameter matrices in MVM. These specifications do not allow for covariance spill-overs and feedbacks, which are aspects of major interest. On the contrary, the structured specifications proposed in this paper allow for covariance spill-overs and feedbacks from neighbors. For instance neighbors can be defined as stocks from the same sector; this allows to interpret coefficients as representing sectorial effects. In this sense the present specifications reflect a factor structure, associated e.g. with the classification 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 proximity structure derived from economic rationale.[2] Structured specifications are hence easier to interpret, because

[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 unfeasible in large cross-sections.

[2] In statistical factor analysis, the literature distinguishes between exploratory analysis, where no identifying

factors are defined a-priori.

Silvennoinen and Teräsvirta (2008, section 2) provide a list of desiderata for an ideal MGARCH specification, which applies to MVM more in general. The ideal specification should be:

1. *scalable*, i.e. estimation should be feasible for increasingly large cross-sections;
2. *flexible* enough to allow for covariance spill-overs and feedbacks;
3. *interpretable*;
4. the estimated conditional covariance matrices should be *positive definite* (p.d.) by construction;
5. it would be nice if the resulting Gaussian likelihood function was *easily optimized numerically*; in particular this requires that the calculation of the inverse and of the determinant of the conditional covariance matrix are fast and numerically stable.

Most of the structured specifications have a moderate number of parameters, which often grows linearly with the cross-section dimension n. This means that the number of parameters grows approximately as cn where c is a given positive number. If one fixes the time-series dimension T, increasing the cross-section dimension n makes the ratio of number of parameters to observations converge to a positive fraction $cn/Tn = c/T < 1$, unlike models plagued by the curse of dimensionality. This makes these models scalable.

Thanks to the fact that proximity is derived from economic rationale, the corresponding parameters have a direct economic interpretation. It is also found that if an unrestricted MVM delivers p.d. covariances, then structured restrictions will not change this property. On the contrary, if positive definiteness is not guaranteed, structured specifications may help to obtain p.d. conditional covariance matrices.[3] Hence structured restrictions help also in association with the desideratum of positive definiteness.

Finally, structured specifications usually make the calculation of the inverse and of the determinant of the conditional covariance matrix easier than in the unrestricted case. In several cases these operations can be performed without the use of matrix inversion routines, hence speeding up computations massively.[4] For some structured specifications, the likelihood calculations can be further separated into separate blocks, hence simplifying the computational complexity.[5] Overall, structured specifications appear to satisfy most – if not all – the optimality criteria put forward by Silvennoinen and Teräsvirta (2008).

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.

[3] This is the case in CCC specifications when the unrestricted estimation of the correlation matrix is not feasible, see Subsection 5.7.

[4] See Subsections 2.4, 5.12.

[5] This is for instance the case in spatial GO-GARCH specifications, see Subsection 5.6.

Spatial statistics has a long tradition in modeling unconditional variance matrices, see e.g. Cressie (1993) and references therein. Recently, Bera and Simlai (2004) and Yan (2006) have considered a spatial cross-section model with a form of heteroskedasticity analogous to ARCH and to stochastic volatility models respectively. In this paper we present a complementary type of spatial volatility model; namely, we introduce spatial specifications within well known time-series models of second moments.

Important recent applications of spatial models in economics can be found in the literature on regional income growth (Riviera-Batiz and Romer 1991), in microeconomics of product diffusion (Brock and Durlauf, 2001), in the analysis of interaction of policy makers in public economics (Brueckner, 2003).

Spatial econometrics has steadily developed over the years, see e.g. Anselin (1988, 2002) and reference therein. Spatial models are used to account for error dependence in regression models, see e.g. Lee (2002), as well as to model individual heterogeneity in panels, see Baltagi et al. (2007), McAleer, Medeiros and Slottje (2008), Pesaran and Tosetti (2008). Despite the wide use of spatial models for unconditional covariance structures, their use in modeling conditional covariances discussed in this paper is novel.

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 (Conley and Ligon 2002, Pesaran et al. 2004) and social proximity (Conley and Topa 2002). Huse (2006) and Gall et al. (2004) report applications of spatial models to the term structure of interest rates (or forward rates), where distance is in terms of maturity.

The simplest notion of proximity is, however, the one inherited from lattice models. Case (1991), for instance, considered this simple notion of proximity, and 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.

As most of the literature, we consider (quasi-) likelihood-based inference. We discuss when this type of inference can be decomposed into subsystem estimation procedures with no loss of information, see Subsection 5.12, where the latter usually involve faster computations. Asymptotic properties of estimators and tests of unrestricted MGARCH processes are discussed in Ling and McAleer (2004), Comte and Lieberman (2003); a complete discussion of how this theory can be adapted to the present class of models is beyond the scope of the paper, and it will be pursued elsewhere.

Many variations of structured MVM specifications exist, and in general the best specification depends on the given application. In order to illustrate how general spatial tools are, we present several structured specifications for major MVM model classes. We find that most models allow a structured specification, although some model classes are more amenable to structured specification than others.

Because there are many possible variations of the same model class, we consider only the simplest representative model from the class. For instance when dealing with Multivariate GARCH (MGARCH) models, we discuss the MGARCH(1,1) model, and observe that extensions for the general MGARCH(p, q) case are very similar to the leading case. The same principle is also employed with respect to asymmetric effects of innovations and other standard variations in ARCH models.

The rest of the paper is organized as follows. Section 2 reviews some basic spatial models; Section 3 introduces weight matrices for asset returns. The setup of MVM models is discussed in Section 4. MGARCH models are discussed in Section 5, and other MVM in Section 6. A discussion on how to define more general, and possibly time-varying, weight matrices is reported in Section 7. Section 8 reports concluding remarks. The Appendix contains definitions and properties of weight and spatial matrices.

2 Spatial covariance models

In this section we review simple spatial processes and their interpretation in terms of weight matrices and the associated neighboring structure. For a more detailed introduction we refer to Cressie (1993) and references therein.

2.1 SAR processes

Consider realizations on m random variables u_i, where each i corresponds to one of m units. We assume that u_i are mean zero, $\mathbb{E}(u_i) = 0$ where $\mathbb{E}(\cdot)$ denotes the expectation operator, and that there exist a proximity structure on the units, represented by the weights w_{ij} which take a positive value if units i and j are (different) neighbors ($i \neq j$), and value 0 otherwise.

The simplest example of spatial process is given by the Spatial AutoRegressive process of order 1, SAR(1), given by

$$u_i = \phi \sum_{j=1}^{m} w_{ij} u_j + \varepsilon_i \tag{1}$$

where $w_{ij} \geq 0$ are proximity weights and ε_i are errors. The vector $\varepsilon := (\varepsilon_1 : \cdots : \varepsilon_m)'$ has expectation 0 and positive definite covariance matrix V; sometimes it is also assumed to be Gaussian. Usually $V := \text{diag}(v)$ where v is an $m \times 1$ vector of positive elements, $v > 0$ and we indicate by $\text{diag}(v)$ is the $m \times m$ diagonal matrix with the elements of v on the main diagonal. Often a scalar specification is employed, i.e. $v = \sigma^2 1_m$, where 1_m is an $m \times 1$ vector with all elements equal to 1 and $\sigma^2 > 0$.

Eq. (1) postulates that u_i is influenced by measurements on (first order) neighbors, whose measurements are picked out by the w_{ij} weights; ε_i is the error in the equation. The coefficient ϕ represents the spatial AR parameter, which is usually restricted to be the same for all i; we call this the 'homogeneity assumption', and it can be relaxed, see Subsection 2.3 below.

Collecting the vector of u_i random variables in the vector $u := (u_1 : \cdots : u_m)'$ and defining

the weight matrix $W := (w_{ij})$, the SAR(1) process can be written as

$$u = \phi W u + \varepsilon \tag{2}$$

The matrix W is $n \times n$, it has 0 across the main diagonal[6], and row i of matrix W, $w_i' := (w_{i1} : \cdots : w_{im})$ has positive elements corresponding to the first order neighbors of unit i. Note that $w_i'u$ is proportional to a (possibly weighted) average of measurements on first-order neighbors of unit i. The matrix W is called 'weight matrix' in the following; here we assume weight matrices to be normalized across the rows, $\sum_{j=1}^{m} w_{ij} = 1$, see the Appendix for a discussion of this normalization. Possible ways to construct weight matrices W are presented in Section 7 below.

A similar expression holds for SAR(2) processes, by defining two weight matrices W_1 and W_2, whose elements in row i define the first order and second order neighbors, respectively, of unit i. The resulting SAR(2) model can be written as $u = \phi_1 W_1 u + \phi_2 W_2 u + \varepsilon$. The general form of a SAR(p) model is

$$u = \sum_{h=1}^{p} \phi_h W_h u + \varepsilon \tag{3}$$

where each W_h is a weight matrix associated with neighbors of order h, each of which has AR parameter ϕ_h.

2.2 Covariance structure and spatial matrices

SAR processes in eq. (2) and (3) are used to model the covariance structure of u; in particular one can solve the equations for u by computing $\left(I - \sum_{h=1}^{p} \phi_h W_h\right) u = \varepsilon$, from which

$$M := \mathbb{V}(u) = S^{-1} V S^{-1\prime}, \tag{4}$$

$$S = I - \sum_{h=1}^{p} \phi_h W_h, \tag{5}$$

where $\mathbb{V}(\cdot)$ is the variance operator. Note that the Gaussian hypothesis is not required to obtain (4).

The covariance structure (4) implied by the SAR model can be quite articulate, despite being generated by just a few coefficients; for the SAR(1) case, one typically has one ϕ parameter and one variance parameter σ^2 in $V = \sigma^2 I$.

2.3 Heterogeneous SAR processes

One generalization of the SAR(1) process is obtained by assuming a different AR coefficient ϕ in (1) for each unit; this results in

$$u_i = \phi_i \sum_{j=1}^{m} w_{ij} u_j + \varepsilon_i,$$

[6] Matrices with 0s across the diagonal are sometimes called 'hollow' matrices.

where ϕ_i indicates the AR coefficient ϕ for unit i. In matrix notation:

$$u = \text{diag}(\phi)Wu + \varepsilon \tag{6}$$

where $\phi := (\phi_1 : \cdots : \phi_m)'$ is the $m \times 1$ vector of AR coefficients. The corresponding generalized SAR(p) process is

$$u = \sum_{h=1}^{p} \Phi_h W_h u + \varepsilon \tag{7}$$

where $\Phi_h := \text{diag}(\phi^h)$, $\phi^h := (\phi_1^h : \cdots : \phi_m^h)'$ and ϕ_i^h is the AR(h) coefficient for unit i. This implies a covariance structure of the type (4) with

$$S := S_0 + \sum_{h=1}^{p} S_h W_h. \tag{8}$$

with $S_0 = I_n$, $S_h := -\Phi_h$. We call (8) the 'heterogeneous' specification. We call a matrix S of the form (8) a *spatial matrix*; in the Appendix we give a formal definition and properties of spatial matrices.

The coefficients ϕ_i^h that appear in (7) are not necessarily all distinct. One possibility that all ϕ_i^h in ϕ^h are equal, reducing the associated process to the case of the standard SAR; this case is called the 'homogeneous' specification. We also consider the case in which ϕ_i^h is equal to ϕ_j^h for all j that are first-order neighbors to unit i. 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.

2.4 Computation of the inverse and the determinant

It is interesting to observe that the inverse of the covariance matrix M implied by a (generalized) SAR process, see eq. (4), can be computed without using matrix-inversion routines. In fact

$$M^{-1} = SV^{-1}S', \qquad V^{-1} = \text{diag}(1_m./v), \tag{9}$$

where $a./b$ indicates element-wise division, $1_m./v = (1/v_1 : \cdots : 1/v_m)'$. Hence the inverse of M is computed by the following steps:

1. compute $S = I - \sum_{h=1}^{p} \Phi_h W_h$, which is a linear function of the parameters in ϕ^h;
2. compute $V^{-1} = \text{diag}(1_m./v)$, which just involves elementwise division;
3. perform the matrix multiplication $SV^{-1}S'$ to obtain M^{-1}.

The computation of this inverse is hence fast and feasible also for large cross-sections; see also Subsection 4.3 below.

The same remarks apply to the inversion of the correlation matrix R corresponding to M. Let $\text{dg}(M)$ be the $m \times 1$ vector containing the diagonal elements in M, and let $R = \text{cor}(M)$ be the correlation matrix obtained by standardizing M, i.e. $R = \text{cor}(M) := D_M^{-\frac{1}{2}} M D_M^{-\frac{1}{2}}$, where $D_M := \text{diag}(\text{dg}M)$. One finds

$$R^{-1} = D_M^{\frac{1}{2}} M^{-1} D_M^{\frac{1}{2}}$$

where $D_M^{\frac{1}{2}} = \text{diag}(\text{dg}M)^{\frac{1}{2}}$ is a diagonal matrix with element i on the main diagonal equal to $\sqrt{M_{ii}}$, the square root of the diagonal elements of M. Hence inversion of R is only slightly more time-consuming than the inversion of M; it consists of adding the following final step to the inversion of M:

4. multiply all elements in row i by $\sqrt{M_{ii}}$ and all elements in column j by $\sqrt{M_{jj}}$ to obtain $R^{-1} = D_M^{\frac{1}{2}} M^{-1} D_M^{\frac{1}{2}}$.

Computation of the determinant of the matrix M or R is similarly simplified by the structure $M = S^{-1}VS^{-1\prime}$. One finds $\det M = \det V / (\det S)^2$. The computation of $\det S$ depends on the weight matrices W_h; we report some examples of these computations in Subsections 4.3 and 5.6 below.

2.5 Other spatial processes

In this subsection we briefly describe two other spatial processes, the Spatial Moving Average (SMA) process and the Conditionally AutoRegressive (CAR) process. These models are introduced to demonstrate the central role of spatial matrices also for these alternative spatial processes.

The SMA(1) process corresponds to $u = \varepsilon + \phi W \varepsilon$, which implies $\mathbb{V}(u) = (I + \phi W) V (I + \phi W)'$. For generic SMA($p$) processes one has

$$\Sigma := \mathbb{V}(u) = SVS', \qquad S := I - \sum_{h=1}^{p} \phi_h W_h,$$

which is similar to (4) except that the spatial matrix S appears in the expression of Σ in place of S^{-1} . Note that the inversion of Σ now involves also the inversion of S, $\Sigma^{-1} = S'^{-1} V^{-1} S^{-1}$. For some specifications of the weight matrices W_h this can be obtained analytically, but this is not true in general.

A CAR(1) process under normality implies the following covariance structure

$$\mathbb{V}(u) = (I - \phi W)^{-1} L$$

where L is diagonal and such that $L^{-1}W$ is symmetric. Note the role of the spatial matrix $I - \phi W$ also for this process.

SMA and CAR processes are similar in nature to SAR processes[7]; for simplicity in the following we will concentrate attention on SAR as a leading example of spatial processes.

2.6 Time series analogues to SAR

In the rest of the paper we use the notions of spatial matrices in order to obtain structured specification for MVM when u has a time subscript t. In particular we employ (4) as a way

[7] For a discussion of relative merits of CAR(1) versus SAR(1) processes see Martellosio (2008), section 2.1.

to specify covariance matrices (or the implied correlation matrices). This corresponds to the following analogues of the generalized SAR process in (7):

$$u_t = \left(\sum_{h=1}^{p} \Phi_h W_h \right) u_t + \varepsilon_t.$$

where $\Phi_h = \text{diag}(\phi^h)$. Here the AR matrix is a spatial matrix, with zeros on the main diagonal.

Moreover, we also use the following, different time-analogue of the generalized SAR process

$$u_t = \left(\sum_{h=0}^{p} \Phi_h W_h \right) u_{t-1} + \varepsilon_t$$

where $W_0 := I_n$. Note that the resulting AR-parameters $S := \sum_{h=0}^{p} \Phi_h W_h$ are spatial matrices of the type (8), with diagonal elements (possibly different from 0) which represent diagonal effects of the own lags.

3 Spatial proximity for asset returns

In this section we give examples of possible weight matrices W to be used in the context of MVM. As a leading example, we consider a set of 12 stock returns over a single time-period, $u := (u_1 : \cdots : u_{12})'$. Because there is a single time period, we do not indicate the time subscript t in u. We return to the full notation in Section 4. We assume that the (conditional) mean of u is 0 and that we wish to model the covariance structure of u as a SAR process.

The following subsections discuss the definition of weight matrices W when distances are defined on the basis of groups, where groups are induced by a (single in Subsection 3.1 or composite in Subsection 3.2) classification criterion. Extensions to higher order neighbors are discussed in the Subsection 3.3. The classification criteria are indicated by the letters F, G, H.

3.1 A single classification criterion

Assume that the 12 assets can be classified on the basis of a single classification criterion, called F. In this example we take F to represent each stock's sector; for simplicity we let F assume only two values, such as 'goods' and 'services', which we label as 1 and 2 respectively. We record in F_i the level of the factor in unit i, which is 1 for 'goods' and 2 for 'services'.

Assume that 6 stocks belong to sector 1 and the other 6 to sector 2, and that the first 6 in u correspond to the first group. We next discuss how to obtain the corresponding W^F weight matrix that defines the first-order neighbors as members of the same group. i.e. stocks from the same sector.

The first-order neighbors of stock 1 are given by stocks 2 to 6, and hence the first row of W^F is given by the row vector $\frac{1}{5}(0 : 1_5' : 0_6')$, where 0_s and 1_s are $s \times 1$ vectors with all elements equal to 0 or 1 respectively. Similarly the second row of W^F is given by $\frac{1}{5}(1 : 0 : 1_4' : 0_6')$ etc.

For stock number 7, one finds that the corresponding row of W^F is given by $\frac{1}{5}(0'_7 : 1'_5)$, for number 8 it is $\frac{1}{5}(0'_6 : 1 : 0 : 1'_4)$ etc.

Grouping these rows together one obtains

$$W^F = \text{diag}(J_6, J_6), \qquad J_s := \frac{1}{s-1}\left(1_s 1'_s - I_s\right)$$

where $\text{diag}(A, B)$ is a block-diagonal matrix with the A and B on the main diagonal. Note that, if the two groups have different number of stocks, n_1 and n_2 say, the diagonal blocks become J_{n_1} and J_{n_2}, $W^F = \text{diag}(J_{n_1}, J_{n_2})$.

This example can be directly extended to the case of n assets, $u := (u_1 : \cdots : u_n)'$, classified according to a classification scheme F with k classes, labeled $1, \ldots, k$.[8] Specifically, let n_h be the number of assets belonging to class h, with $n = \sum_{h=1}^{k_1} n_h$. Assume, without loss of generality, that the first n_1 assets in the vector u refer to the stocks in class 1, the second group of n_2 to class 2 and so forth. It is simple to verify that the associated weight matrix of first-order neighbors is of the type

$$W^F = \text{diag}(J_{n_1}, \ldots, J_{n_{k_1}}). \tag{10}$$

The specification of the W^F matrix can be obtained in general terms as follows. Let $w_{ij} := (W^F)_{ij}$; one has

$$w_{ij}^F = \frac{w_{ij}^*}{\sum_{j=1}^n w_{ij}^*}, \qquad w_{ij}^* := 1\left(F_i = F_j, i \neq j\right), \tag{11}$$

where 1(condition) is the indicator function that takes value 1 if the condition is true and value 0 otherwise. The format (11) illustrates that w_{ij} are normalized weights along each row, and that $w_{ij} \neq 0$ when $F_i = F_j$ and $i \neq j$.

3.2 More than one classification criterion

In this subsection we discuss how to construct weight matrices in case more than one classification criterion is present. The idea is that one can apply the same principles and techniques that apply in ANalysis Of VAriance, ANOVA, see e.g. Wichura (2006).

Consider again the 12 asset example and a second classification criterion G, which can be taken to be capitalization size. Assume also that G has three levels, 'large', 'medium' and 'small', labeled 1, 2, 3 respectively. The factor level of unit i is indicated as G_i. Assume in the example that $G_i = 1$ for stocks numbered i =1, 2, 7, 8, $G_i = 2$ for i = 3, 4, 9, 10 and $G_i = 3$ for i =5, 6, 11, 12, see Table 1.

If G were the only classification criterion, the previous subsection indicates how to construct the first-order-neighbors weight matrix. Specifically, let $u^* = Pu$ be a re-ordering of the elements in u which moves all elements i with in $G_i = 1$ as the first group in u^*, the elements with $G_i = 2$ as the second group in u^* and the ones with $G_i = 3$ as the final group in u^*. Note that the matrix P is a permutation matrix and that for instance, $u^* := (u_1 : u_2 : u_7 : u_8 : u_3 : u_4 : u_9 : u_{10} : u_5 : u_6 : u_{11} : u_{12})'$.

[8] Obviously, these classes form a classification scheme if they are disjoint and they cover all possible cases.

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

Table 1: The table contains label-numbers of stocks in the example with 12 assets returns, with a classification induced by two factors F and G.

Then, the weight matrix for u^* is $W_2^* = \text{diag}(J_4, J_4, J_4)$ and it is simple to see that the corresponding weight matrix for u is

$$W^G = PW_2^{*\prime}P = P\text{diag}(J_4, J_4, J_4)P', \tag{12}$$

which has the same structure as for the case of F, except for the presence of the permutation matrix P.

One possible combination of the factors F and G can be obtained by considering a spatial matrix $S = \phi_F W^F + \phi_G W^G$, where W^F is defined in (10) and W^G in (12). Here the effect of both factors are additive. This gives a simple way to combine the effects of the two classification factors F and G.

A more general combination of factors F and G can be obtained by considering each cell in the Table 1 as a group; this allows to measure interactions between the two factors. Specifically let H be the classification obtained by considering the Cartesian product of elements in F and G, i.e. 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. Specifically for the 12-returns example, 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. A discussion on the number of classification factors is reported in the Subsection 3.3.

3.3 Second order neighbors and the number of factors

The above discussion focused on first-order neighbors. In this subsection we discuss higher-order neighbors and the choice of number of classification factors.

Higher-order neighbors can be included in the specification of spatial matrices S by adding a corresponding weight matrix. For a factor F corresponding to 'industry', one may define second order neighbors as members of other industries in the same sector, for instance. This corresponds to matrices W_1^F and W_2^F, and the discussion in the previous subsections applies. Note that for some order, some unit j may have no neighbors, which results in a 0 row in the corresponding W matrix. For some even higher order, one may have that all units have no neighbor, and this gives a maximum for the neighboring order.

Assume that a weight matrix W has row i equal to 0; the corresponding coefficient ϕ is potentially (generically) identified in the homogeneous case (5). In the heterogeneous case (8) the parameters ϕ_i^h is not identified. This shows that identification of the ϕ^h vectors in (8) requires care.

We next discuss the choice of classification factors. Too many factors may induce cells in the Cartesian product that have 1 unit; this would result in a 0 row for the associated weight matrix, which contradicts the assumption that all row sums to 1, $W1 = 1$. This reflects the idea that spatial models are meaningful if each unit has at least one neighbor. However, zero rows can be accommodated in products like $\text{diag}(\phi)W$ by choosing the corresponding elements in ϕ to equal 0, see the Appendix.

A different limitation to the specification of the spatial matrix S in (8) is given by the fact that an unrestricted S has m^2 entries, and the spatial specification is meant to reduce this number. If for instance, the distinct elements in Σ are identified, this means that there are at most $m(m+1)/2$ identifiable coefficients in S and V. This means that the number of different coefficients ϕ_i^h in (8) must be (greatly) less that this upper limit.

We finally observe that identification problems in $S = S_1W_1 + S_2W_2$ arise when (a subset of the rows of) two weight matrices W_1 and W_2 are identical, a situation which corresponds to the case where neighbors implied by the corresponding factors are identical. Take for simplicity $S_1 = \text{diag}((\phi_1 1_6' : \phi_2 1_6')')$, $S_2 = \text{diag}((\phi_3 1_6' : \phi_4 1_6')')$ and assume that $(I_6 : 0)W_1 = (I_6 : 0)W_2$, i.e. that the first block of 6 rows in W_1 and W_2 are equal. Then

$$(I_6 : 0)S = (\phi_1 + \phi_3)\,(I_6 : 0)W_1$$

depends on $\phi_1 + \phi_3$, and one cannot identify ϕ_1 and ϕ_3 separately.

The discussion of this identification problem is difficult in general terms, because it depends on the specific set of weight matrices W; this should be checked on a case-by-case basis. In general, it is wise to consider only a small set of classification factors, as well as to try to minimize the number of different parameters in the heterogenous parametrization. In the following exposition, we adopt a single classification criterion for simplicity.

4 Spatial MVM setup

In this section we discuss the definition of weight matrices and spatial matrices for MVM. We consider the case of a single proximity criteria and first-order neighbors. Extensions to higher-order models can be carried out along the lines of the previous section, and they are not discussed here.

4.1 Conditional covariances

In the following, we consider a set of returns $y_t := (y_{1,t} : \cdots : y_{n,t})'$ on n assets, classified into k groups with the use of some economic or financial classification criterion. The assets within each group are taken as first-order neighbors. The k sets have dimension n_1, n_2, ..., n_k with

$\sum_{i=1}^{k} = n$, and the assets are ordered according to these groups; in other words $y_{1,t}, \ldots, y_{n_1,t}$ refer to assets in first group, $y_{n_1+1,t}, \ldots, y_{n_1+n_2,t}$ refer to assets in the second group, and so on.

We 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. We assume that the conditional mean of y_t is 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 $\mathbb{E}_{t-1}(u_t) = 0$; interest lies in the prediction of

$$\Sigma_t := \mathbb{V}_{t-1}(u_t) = \mathbb{E}_{t-1}(u_t u_t').$$

4.2 Weight matrices

Before discussing MVM models, we first describe the prototype weight matrices encountered in this context. The weight matrix corresponding to the classification of the elements in y_t or u_t is given by

$$W_n := \text{diag}(J_{n_1}, \ldots, J_{n_{k_1}}), \tag{13}$$

see (10). Here W_n is $n \times n$.

When considering vec (Σ_t) or vec $(u_t u_t')$, element number h in these vectors corresponds to a pair of units (i, j), where $h = (i-1)n + j$ and vec is the column-stacking operator. Note that $(\Sigma_t)_{ij}$ is the conditional covariance between assets i and j. The same elements appear in vech (Σ_t), where vech stacks the elements on and below the main diagonal; we describe the weight matrices for vec (Σ_t) and show how to adapt these results for vech (Σ_t) specifications.

In order to construct weight matrices for the elements in vec (Σ_t), one needs to define first-order neighbors for pairs (i, j) and (l, m), say. Let (i, j) correspond to element $h = (i-1)n+j$ in vec (Σ_t) and (l, m) correspond to element $v = (l-1)n+m$. One can define a weight matrix that classifies as first order neighbors of element h, corresponding to the pair (i, j), all elements v, corresponding to the pair (l, m) for which $(F_i = F_l, F_j = F_m)$, i.e. for which the sectors coincide. Moreover one wishes to exclude from v the pairs for which $i = l$ or $j = m$, i.e. where one of the stocks is repeated.

In other words, when modeling the conditional covariance $(\Sigma_t)_{ij} = \mathbb{E}_{t-1}(u_i u_j)$ of excess returns on assets i and j, first order neighbors correspond to all other asset pairs (l, m) where the first asset l is in the same sector of asset i and the second asset m is in the same sector as asset j, without counting pairs where $i = l$ or $j = m$.

This results in a weight matrix with (h, v) element w_{hv} of the form

$$w_{hv} = \frac{w^*_{hv}}{\sum_{v=1}^{n^2} w^*_{hv}}, \qquad w^*_{hv} := 1(F_i = F_l, F_j = F_m, i \neq l, j \neq m), \tag{14}$$

where we have used an expression similar to (11). We label the resulting matrix as $W_{n^2,3}$.

An alternative weight matrix is obtained by replacing w^*_{hv} in (14) with the following alternative choices:

$$w^{*(1)}_{hv} := 1\left(F_i = F_l, i \neq l, j = m\right), \qquad w^{*(2)}_{hv} := 1\left(F_j = F_m, i = l, j \neq m\right). \tag{15}$$

We label the associated weight matrices as $W_{n^2,1}$ and $W_{n^2,2}$ respectively. For the weight matrix $W_{n^2,1}$, the first order neighbors of (i,j) correspond to pairs (i,m) with j and m indicating distinct assets from the same sector. For the weight matrix $W_{n^2,2}$, the first order neighbors of (i,j) correspond to pairs (l,j) with i and l indicating distinct assets from the same sector.

$W_{n^2,j}$, $j=1,2,3$ are representative matrices of the class of weight matrices of dimension n^2. In the example which follow, in order to simplify exposition we use a single weight matrix, labeled W_{n^2}, which can be taken to be one the matrices $W_{n^2,j}$, $j=1,2,3$.

It may be of interest to note the following property; all proofs are placed in the Appendix.

Proposition 1 *In the previous notation*

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

The proposition shows that these matrices have a Kronecker product structure; Proposition 4 in the Appendix discusses necessary and sufficient conditions on H, K in a product $H \otimes K$ to ensure that $H \otimes K$ is a weight matrix.

4.3 Spatial matrices

In this subsection we consider spatial matrices S generated by the weight matrices $W = W_n$ or W_{n^2} described earlier; they take of the form

$$S = S_0 + S_1 W, \qquad S_j = \operatorname{diag}(s^j), \quad j = 0, 1. \tag{16}$$

The heterogeneous specification is obtained when the elements in s^j are all unrestricted; the group-homogeneous case corresponds to assuming $s_i^j = \phi_{F_i}^j$, where $\phi_1^j, \ldots, \phi_k^j$ are group-specific coefficients. Finally the homogeneous case is obtained by imposing $\phi_l^j = \phi_m^j$, $l, m = 1, \ldots, k$.

We also use spatial specifications to model an $n \times n$ correlation matrices R; in particular one can assume that R is the correlation matrix corresponding to a covariance matrix of a SAR process, see eq. (4); in other words we assume

$$R = \operatorname{cor}(S^{-1} V S^{-1\prime}), \qquad V = \operatorname{diag}(v) \tag{17}$$

and where S is parametrized as in (16) with $W = W_n$. Correlation matrices constructed in this way are positive definite (provided V is) and all parameter values are valid for elements in s^j.

We now present an example on the computation of $\det S$ which is required in the calculation of $\det M$, see Subsection 2.4. Here W is assumed to be of the form (13) and s^j in (16) is partitioned conformably into $s^j = ((s_1^j)' : \cdots : (s_k^j)')'$ where each s_h^j is $n_h \times 1$, $h = 1, \ldots, k$. We find

$$\det S = \prod_{h=1}^{k} \det\left(\operatorname{diag}(s_h^0) + \operatorname{diag}(s_h^1) J_{n_h}\right)$$

which shows that the computational problem of finding the determinant of an $n \times n$ matrix is decomposed in k smaller tasks of finding the determinant of $n_h \times n_h$ matrices $\operatorname{diag}(s_h^0) +$

$\text{diag}(s_h^0)J_{n_h}$. This decomposition can be useful also when one wishes to decompose the estimation of the complete systems into subsystems, see the example in Subsection 5.6 below.

Moreover, for the group-homogeneous case one has further simplifications. In fact when $s_h^j = \phi_h^j 1_{n_h}$, $j = 0, 1$ each of the terms $\det(\text{diag}(s_h^0) + \text{diag}(s_h^0)J_{n_i})$ reduces further as follows:

$$\begin{aligned}\varphi(\phi_h^0, \phi_h^1, n_h) &:= \det\left(\text{diag}(s_h^0) + \text{diag}(s_h^1)J_{n_h}\right) = \det\left(\phi_h^0 I_{n_h} + \phi_h^1 J_{n_h}\right) \\ &= \det\left(c_{h0}P_{h0} + c_{h1}P_{h1}\right) = \left(\phi_h^0 - \phi_h^1\right)\left(\phi_h^0 - (n_h - 1)^{-1}\phi_h^1\right)^{n_h - 1},\end{aligned}$$

where $P_{h1} := n_h^{-1} 1_{n_h} 1_{n_h}'$ and $P_{h0} := I_{n_h} - P_{h1}$ are orthogonal projection matrices of rank 1 and $n_h - 1$ respectively and $c_{h0} = \phi_h^0 - (n_h - 1)^{-1}\phi_h^1$, $c_{h1} = \phi_h^0 - \phi_h^1$. Here have used well-known results on eigenvalues of linear combinations of projectors, see Magnus (1982, Lemma 2.1.ii); see also Magnus and Müris (2008) for a recent application of these properties in the specification of covariance matrices for panel data models.[9]

We illustrate the effects of this decomposition in Subsection 5.6 below.

5 Multivariate GARCH models

In this section we review some of the most common MGARCH models; we show how spatial concepts can be used to control the number of parameters. We use the weight matrices introduced in the previous section. In the following we do not comment on restrictions on parameters that guarantee second-order stationarity for conciseness.

The reference pseudo log-likelihood is a Gaussian one, proportional to $\log L = \sum_{t=1}^{T} \ell_t$ with

$$-2\ell_t = \log\det\Sigma_t + u_t'\Sigma_t^{-1}u_t.$$

Here we discuss some popular MGARCH specifications, including the VEC and BEKK, GO-GARCH, CCC, DCC, VCC models. We discuss them in turn, starting from the simplest models. Here we do not attempt to obtain a complete list of applications, but rather to show the potential of applications of structured specification in MGARCH models. An example of likelihood decomposition is given in Subsection 5.12 and a final Subsection comments on some other MGARCH models where structured specification are more difficult to define.

5.1 BEKK

We consider the following representative BEKK specification, see Engle and Kroner (1995):

$$\Sigma_t = C + Au_{t-1}u_{t-1}'A' + B\Sigma_{t-1}B', \tag{18}$$

where C is an $n \times n$ positive definite matrix, A, B are unrestricted $n \times n$ matrices; A, B and C are the parameter matrices to be estimated. See also McAleer et al. (2008) for a motivation of the BEKK specification based on random coefficient autoregressions. This

[9] Obviously, a similar simplification applies for the homogeneous specification when all elements of S_0 and S_1 are identical, $\phi_h^j = \phi^j$, $j = 0, 1$, but this would give cross-subsystem restrictions among the parameters, hence violating the condition that parameters are variation-free.

n	a	b	c	d	e	f	g
	$1000n$	$2.5n^2$	$0.5n^4$	$6n$	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-section dimension; column a: number of data points for time series with $T = 1000$; b, c, d: order of the number of parameters $O(n^2)$, $O(n^4)$, $O(n)$; e, f, g: ratio of number of parameters to sample size for b, c, d; if >1, model is not estimable (more parameters than observations).

specification ensures positive definiteness of Σ_t. The total number of coefficients in (18) is $0.5n\,(n+1) + 2n^2 = 2.5n^2 + O(n) = O\left(n^2\right)$. This number grows proportionally to the square of n, and it becomes unfeasible for estimation in large cross-sections, see columns b and e in Table 2.

In order to economize on parameters, the diagonal-BEKK specification restricts $A = \text{diag}(a)$ and $B = \text{diag}(b)$, $a := (a_1 : \cdots : a_n)'$, $b = (b_1 : \cdots : b_n)'$, and the number of parameters reduces to $0.5n\,(n+1) + 2n = 0.5n^2 + O(n) = O\left(n^2\right)$, which however still grows as n^2 due to the estimation of C.

Note that the diagonal specification does not allow covariance spill-overs; in fact

$$(\Sigma_t)_{ij} = C_{ij} + a_i a_j u_{i,t-1} u_{j,t-1} + b_i b_j \left(\Sigma_{t-1}\right)_{ij},$$

so that each conditional covariance $(\Sigma_t)_{ij}$ depends on its own past, with innovations stemming only from the the corresponding pair $(u_{i,t-1}, u_{j,t-1})$. Hence no covariance spill-overs are possible, and this prevents some needed flexibility.

An even more restricted specification is a scalar one, where $a = \alpha 1_n$, $b = \beta 1_n$ which reduces the number of parameters in A and B but does not affect the rate of increase with n, which is still $0.5n^2 + O(n) = O\left(n^2\right)$ due to the estimation of C. Note that covariance targeting, see e.g. Caporin and McAleer (2008), does not change the number of parameters.

5.2 Structured BEKK

A structured BEKK specification can be obtained by setting $C = S^{-1} V S^{-1\prime}$ and assuming A, B and S to be spatial matrices of the type (8), i.e.

$$A = A_0 + A_1 W_n, \qquad B = B_0 + B_1 W_n, \qquad S = I - S_1 W_n, \tag{19}$$

where $A_j := \text{diag}(a^j)$, $B_j := \text{diag}(b^j)$, $j = 0, 1$, $S_1 := \text{diag}(s^1)$, $V := \text{diag}(v)$ are all diagonal $n \times n$ matrices. The number of parameters is $6n = O(n)$, which grows linearly with n and it is estimable for large cross-sections, see columns d and g in Table 2.

The structured BEKK specification still delivers p.d. matrices provided A, B, S are of full rank. Despite a moderate number of parameters, and unlike the diagonal BEKK, the structured BEKK specification allows for covariance spill-overs. In fact consider the i-th element in

$$v_t := Au_{t-1} = A_0 u_{t-1} + A_1 W_n u_{t-1} = \text{diag}(a^0)u_{t-1} + \text{diag}(a^1)W_n u_{t-1},$$

and observe that v_{it} contains two terms: the first one ($a_i^0 u_{i,t-1}$) contains the own-lag u_{it} term, while the second one contains $a_i^1 w_i' u_{t-1}$, 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 the average of u_{t-1} for stocks in the same sector; hence $w_i' u_{i,t-1}$ represents spill-overs from other stocks in the sector of unit i. Thus the term $Au_{t-1}u_{t-1}'A'$ contains both diagonal effects and spill-over effects from the same sector.

A similar interpretation applies to the $B\Sigma_{t-1}B'$ term in (18). In fact let e_i be the i-th column of I_n; then

$$\begin{aligned}\left(B\Sigma_{t-1}B'\right)_{ij} &= \left(b_i^0 e_i' + b_i^1 w_i'\right)\Sigma_{t-1}\left(b_j^0 e_j + b_j^1 w_j\right) = \\ &= b_i^0 b_j^0 \left(\Sigma_{t-1}\right)_{ij} + b_j^0 b_i^1 \left(w_i'\Sigma_{t-1}e_j\right) + b_i^0 b_j^1 \left(e_i'\Sigma_{t-1}w_j\right) + b_i^1 b_j^1 \left(w_i'\Sigma_{t-1}w_j\right).\end{aligned}$$

The four terms on the r.h.s. have the following interpretation. The first term contains the last value 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 sector F_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 from sectors F_i and F_j (excluding asset i and j). This breakdown clarifies covariance feedback from assets in sectors F_i and F_j onto $(\Sigma_t)_{ij}$; the first term represents a diagonal effect, and the last three are feedback effects from the same sector.

Overall, structured specifications allow to model spill-overs through the innovations and feedback through past conditional covariances from assets belonging to the same sectors. This shows the flexibility of the structured specification, which can accommodate these effects without parameter proliferation.

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 a^j, b^j, s^j, $j = 0, 1$. One option is to restrict $b^1 = 0$, so as to allow covariance spill-overs only through the term $Au_{t-1}u_{t-1}'A'$. Another option is to restrict $b^1 = \beta 1_m$, so as to have a homogeneity for the covariance spill-overs from the term $B\Sigma_{t-1}B'$. An intermediate option is obtained by assuming group-homogeneity for b^1. Similar arguments can be applied to restrictions on b^0, a^1, a^0. Obviously, many sub-models can be constructed by combining restrictions of this type to a selection of a^j, b^j, s^j, $j = 0, 1$.

5.3 VEC

Consider the representative VEC specification, given by

$$\text{vec}\,\Sigma_t = \text{vec}\,C + A\text{vec}\left(u_{t-1}u_{t-1}'\right) + B\text{vec}\,\Sigma_{t-1} \tag{20}$$

where C is an $n \times n$ positive definite matrix, and A and B are $n^2 \times n^2$ parameter matrices, see Bollerslev, Engle and Wooldridge (1988) and Engle and Kroner (1995). The VEC specification does not guarantee that Σ_t is positive definite.

Because of symmetry in Σ_t and $u_t u_t'$, the system (20) can be summarized by using the vech operator and the transformations rules $\text{vec}(\Sigma_t) = D_n \text{vech}(\Sigma_t)$ and $L_n \text{vec}(\Sigma_t) = \text{vech}(\Sigma_t)$, see Magnus and Neudecker (2007) section 3.8. D_n is the $n^2 \times \frac{1}{2}n(n+1)$ duplication matrix and L_n is the $\frac{1}{2}n(n+1) \times n^2$ elimination matrix that extracts the un-repeated elements, see Magnus (1988). The VEC specification can be expressed in terms of vech as follows:

$$\text{vech}\,\Sigma_t = \text{vech}\,C + A^* \text{vech}\left(u_{t-1}u_{t-1}'\right) + B^* \text{vech}\,\Sigma_{t-1} \tag{21}$$

$$A^* = L_n A D_n \qquad B^* = L_n B D_n, \qquad A = D_n A^* L_n, \qquad B = D_n B^* L_n \tag{22}$$

C is a positive definite $n \times n$ matrix, and A^* and B^* are $\frac{1}{2}n(n+1) \times \frac{1}{2}n(n+1)$ unrestricted parameter matrices. This implies that the number of free parameters in A (respectively B) is equal to the number of elements in A^* (respectively B^*), i.e. $\left(\frac{1}{2}n(n+1)\right)^2$. This gives $\frac{1}{2}n(n+1)$ parameters in C and $2\left(\frac{1}{2}n(n+1)\right)^2 = \frac{1}{2}n^4 + O(n^3) = O\left(n^4\right)$ parameters in A and B. Even with relatively small values of n, model estimation becomes soon unfeasible, see columns c and f in Table 2.

In order to restrict the number of parameters, one popular option is to assume a diagonal specification for the matrices A^* and B^*; this gives $\frac{3}{2}n(n+1) = 1.5n^2 + O(n) = O\left(n^2\right)$ coefficients, a number that still increases with the square of n, see columns b and e in Table 2. Hence also the diagonal-VEC is not estimable on large cross-sections. Moreover it also excludes covariance spill-overs, just as the diagonal-BEKK specification, hence also falling short of the requirement of flexibility.

5.4 Structured VEC

A structured VEC specification can be obtained by setting $C = S^{-1}VS^{-1\prime}$ and assuming A, B in (22) and S to be spatial matrices of the type (8), i.e.

$$A = A_0 + A_1 W_{n^2}, \qquad B = B_0 + B_1 W_{n^2}, \qquad S = I - S_1 W_n, \tag{23}$$

where $A_j := \text{diag}(a^j)$, $B_j := \text{diag}(b^j)$, $i = 0, 1$, $S_1 := \text{diag}(s^1)$, $V := \text{diag}(v)$. Here W_n is defined in (13) and W_{n^2} is defined in (14); A_j and B_j are diagonal $n^2 \times n^2$ matrices, while S_i are $n \times n$. The number of unrestricted coefficients in a^j and b^j are $\frac{1}{2}n(n+1)$, because of the duplication in (22). The number of parameters in (23) for the heterogenous specification is $2n(n+1) + 2n = 2n^2 + O(n) = O(n^2)$, which grows like n^2. Hence the heterogeneous structured-VEC specification, while improving the growth rate of number of parameters from $O(n^4)$ to $O(n^2)$, is still over-parametrized, in the sense that its estimation becomes unfeasible for large cross-sections, see columns b and e in Table 2.

When all coefficients in A_j and B_j are restricted to be group-homogeneous, the number of parameters in a^j and b^j becomes k, so that the total number of parameters for the group-homogeneous specification (23) decreases to $4k + 2n = 2n + O(1) = O(n)$, which is linear

in n. If the structured specification is restricted to be completely homogeneous, then the order of the number of parameters is reduced to $4 + 2n = 2n + O(1) = O(n)$, which is of the same order as the group-homogeneous specification. Thus both the group-homogeneous and the homogeneous specifications are estimable on large cross-sections, see columns d and g in Table 2.

Just like the structured-BEKK specification, the structured-VEC specification allows for covariance spill-overs, both in the homogeneous and heterogeneous cases. As for structured-BEKK specifications, the parameter matrices A_0 and B_0 represent 'diagonal effects' while A_1 and B_1 contain the coefficients linking each covariance to covariances of assets belonging to the same groups. In the last case, A_1 includes the spill-over effects of innovations of other assets, while B_1 represents the feedback effects from variances and covariances of first order neighbors.

While imposing a homogeneous ARCH or GARCH dynamic across all assets may be deemed too restrictive, the assumption of equal ARCH/GARCH dynamics for assets grouped on the basis of economic or financial criteria may be reasonable. The use of groups in the parametrization of the conditional variances has already been used by Billio et al. (2005), Billio and Caporin (2007), Asai et al. (2008) and Bonato et al. (2008).

As evidenced in Engle and Kroner (1995), the BEKK specification is a special case of the VEC specification. Here we show that this nesting remains true for the structured-BEKK and structured-VEC specifications, for an appropriate choice of the weight matrices. In fact, taking vec of equation (18), one finds

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

where we used $\operatorname{vec}(ABC) = (C' \otimes A)\operatorname{vec} B$. The matrices $A \otimes A$ and $B \otimes B$ are spatial matrices of the type (8) with respect to the set of weight matrices $\mathbb{W} := \{W_{n^2,j}, j = 1, 2, 3\}$ see Proposition 1 in Subsection 4.2 and Proposition 6 in the Appendix. Hence the structured-BEKK specification is nested within a structured-VEC specification corresponding to an extension of (23).

5.5 GO-GARCH

We consider the GO-GARCH specification proposed by van der Veide (2002) as an extension of the Orthogonal GARCH of Alexander (2001); the covariance matrix Σ_t is decomposed into

$$\Sigma_t = XV_tX', \qquad V_t = \operatorname{diag}(v_t) \tag{24}$$

where X is an $n \times n$ nonsingular matrix; the dynamics of the conditional variances v_t is defined as univariate GARCH processes of the type $v_{it} = (1 - \alpha_i - \beta_i) + \alpha_i u_{it-1}^2 + \beta_i v_{it-1}$, or, in matrix notation,

$$v_t = c + A\,(u_{t-1} \odot u_{t-1}) + Bv_{t-1}. \tag{25}$$

Here $\odot$ is Hadamard's element-wise product and $u_{t-1} \odot u_{t-1}$ contains the squares of the elements in u_{t-1}; moreover $c = (I - A - B)1_n$ is a variance-targeting vector of constants,

$A := \text{diag}\,(a)$, $B := \text{diag}\,(b)$, $a = (\alpha_1 : \cdots : \alpha_n)'$, $b = (\beta_1 : \cdots : \beta_n)'$. a, b are $n \times 1$ unrestricted parameter vectors to be estimated along with X.

This specification ensures positive definiteness of Σ_t when elements in a and b are non-negative. The total number of coefficients is $n^2 + 2n = n^2 + O(n) = O\left(n^2\right)$. This number grows proportionally to the square of n, and it becomes unfeasible for estimation in large cross-sections, see columns b and e in Table 2.

5.6 Structured GO-GARCH

A structured GO-GARCH specification can be obtained by assuming that the conditional covariance matrix (24) to be an instance of a SAR covariance structure $\Sigma_t = S^{-1}V_t S^{-1\prime}$, see eq. (4), i.e. $X = S^{-1}$ in (24). This assumption corresponds to a SAR process of the form

$$u_t = \Phi W_n u_t + \varepsilon_t$$

where $\mathbb{V}_{t-1}(\varepsilon_t) = V_t = \text{diag}(v_t)$. Here $S = I - \Phi W_n$, $\Phi = \text{diag}(\phi)$. The SAR equations can be interpreted as showing the contemporaneous effect exerted on unit i from first order neighbors. The shocks ε_t represent structural shocks, in the sense of uncorrelated ones, with dynamics in the conditional variances governed by (25).[10]

The matrices A, B in (25) and S can be assumed to be of the type (8). This gives

$$A = A_0 + A_1 W_n, \qquad B = B_0 + B_1 W_n, \qquad S = I_n - \Phi W_n, \tag{26}$$

where $A_j := \text{diag}(a^j)$, $B_j := \text{diag}(b^j)$, $i = 0, 1$ and $\Phi := \text{diag}(\phi)$. Here W_n is defined in (13). The vector c may be left unrestricted or in can be constrained for variance targeting. The parameters in a^j and b^j can be restricted to be nonnegative, in order to guarantee positive conditional variances in (25). The interpretation of parameters is similar to that of the VEC model: the coefficients in A_1 and B_1 represent, respectively, the spill-over and feedback effects of first-order neighbors. They are restricted to 0 in the original GO-GARCH specification.

The number of unrestricted coefficients in a^j, b^j, ϕ, c is $6n = O(n)$, which is linear in n. Group-homogeneous specifications can be employed to reduce the number of parameters. Hence structured-CCC specifications can be estimated also for large cross-sections, see columns d and g in Table 2.

The inversion of Σ_t is simple for structured-GO-GARCH specifications, because $\Sigma_t^{-1} = S' V_t^{-1} S$ where $V_t^{-1} = \text{diag}\,(1_n ./ v_t)$, see Subsection 2.4. Analogously one finds $\log\det\Sigma_t = \log\det V_t - 2\log\det S$ where $\log\det V_t = \sum_{i=1}^{k} \log v_{it}$ and $\det S$ can be calculated as in Subsection 4.3. In Subsection 5.12 below we show how the computations can be further reduced in case of a group-homogeneous specification. Hence structured-GO-GARCH specifications appear to satisfy also the requirement of fast computations of the ideal specification.

[10] The sources of innovations in the GARCH dynamics could be alternatively taken to be $\varepsilon_{t-1} \odot \varepsilon_{t-1}$ instead of $u_{t-1} \odot u_{t-1}$ in (25).

5.7 CCC

The CCC specification by Bollerslev (1990) and Ling and McAleer (2003), decomposes the covariance matrix as

$$\Sigma_t = D_t R_t D_t \qquad D_t = \operatorname{diag}(h_t) \tag{27}$$

where h_t is a $n \times 1$ vector of conditional standard deviations and R_t is a correlation matrix. $R_t = R$ is assumed to be time-invariant and the dynamics of the conditional variances $v_t := h_t \odot h_t$ is assumed to be of the form

$$v_t = c + A(u_{t-1} \odot u_{t-1}) + Bv_{t-1} \tag{28}$$

where c is a $n \times 1$ vector (of positive constants) and A and B are square $n \times n$ unrestricted parameter matrices. R is a positive definite matrix with ones along the main diagonal (i.e. a correlation matrix). This gives n parameters in c and $2n^2$ parameters in A and B and $\frac{1}{2}n(n-1)$ parameters in R. The total number of parameters is hence $2.5n^2 + O(n) = O(n^2)$.

The original CCC specification proposed by Bollerslev (1990) restricts A and B to be diagonal, reducing the number of coefficients in (28) to $3n + \frac{1}{2}n(n-1) = 0.5n^2 + O(n) = O(n^2)$, which is still of the same order of magnitude in n. Hence model estimation is feasible only for small to moderately-sized cross-sections, see columns b and e in Table 2.[11]

5.8 Structured CCC

A structured CCC specification can be obtained by assuming A, B in (28) to be spatial matrices of the type (8) and $R = \operatorname{cor}(S^{-1}VS^{-1\prime})$, to be a correlation matrix of the type (17), i.e.

$$A = A_0 + A_1 W_n, \qquad B = B_0 + B_1 W_n, \qquad S = I_n - \Phi W_n, \tag{29}$$

where $A_j := \operatorname{diag}(a^j)$, $B_j := \operatorname{diag}(b^j)$, $i = 0, 1$, $\Phi := \operatorname{diag}(\phi)$, $V := \operatorname{diag}(v)$. Here W_n is defined in (13).

This specification can be described in terms of a SAR process as follows. Define $u_t^* = G_t u_t$, where $G_t = \operatorname{diag}(g_t)$ is any $\mathcal{F}_{t-1}$-measurable, square scale matrix of full rank, so that u_t^* and u_t have the same conditional correlation matrix. Assume that u_t^* follows a SAR process of the form

$$u_t^* = \Phi W_n u_t^* + \varepsilon_t$$

where $V := \mathbb{V}_{t-1}(\varepsilon_t) = \operatorname{diag}(v)$. This implies that $\mathbb{V}_{t-1}(u_t^*) = S^{-1}VS^{-1\prime}$, $S := I - \Phi W_n$ and hence $R = \operatorname{cor}(S^{-1}VS^{-1\prime})$ is the conditional correlation matrix both for u_t and u_t^*. The SAR equations can be interpreted as showing the contemporaneous effect exerted on u_t^* from first order neighbors $W_n u_t^*$. The shocks ε_t represent structural shocks, in the sense of uncorrelated ones, assumed to have time-invariant second moments.

The parameters in a^j and b^j can be restricted to be nonnegative, in order to guarantee positive conditional standard deviations in (28). The interpretation of parameters is similar

[11] In order to obtain a positive definite estimator of R one simply needs $T > n$. However for $T - n$ small, there is a large variance of the estimated correlation matrix, rendering correlations highly unstable.

to that of the VEC model: the coefficients in A_1 and B_1 represent, respectively, the spill-over and feedback effects of first-order neighbors.

The number of unrestricted coefficients in a^j, b^j, ϕ, v are $6n = O(n)$, which is linear in n. Group-homogeneous specifications can be employed to reduce the number of parameters. Hence structured-CCC specifications can be estimated also for large cross-sections, see columns d and g in Table 2.

The inversion of Σ_t is simple for structured-CCC specification, because $\Sigma_t^{-1} = D_t^{-1} R_t^{-1} D_t^{-1}$ where $D_t^{-1} = \text{diag}\,(1_n./h_t)$ and $R_t^{-1} = R^{-1}$ is computed as detailed in Section 2.4. Hence the structured-CCC specification satisfies the requirement of fast computations of the ideal specification in the Introduction.

Similar structured specifications can be entertained for the Double Smooth Transition Conditional Correlation of Silvennoinen and Teräsvirta (2007), which generalizes CCC by allowing R to smoothly change between two values.

5.9 DCC

We here describe the DCC model of Engle (2002); see also Ding and Engle (2001), Cappiello, Engle and Sheppard (2006), Pelletier (2006), Aielli (2008) for extensions and modifications. As in the CCC class, the conditional covariance is decomposed as in (27) and the conditional correlation matrix R_t is assumed to be the correlation matrix of a $n \times n$ process Q_t, $R_t = \text{cor}\,(Q_t)$, with dynamics

$$Q_t = (1 - \alpha - \beta)\, R + \alpha u_{t-1}^{\diamondsuit} u_{t-1}^{\diamondsuit\prime} + \beta Q_{t-1} \tag{30}$$

where $u_t^{\diamondsuit} := \text{diag}(\text{dg}(Q_t))^{\frac{1}{2}} D_t^{-1} u_t$, see Aielli (2008). R is a positive definite matrix with ones along the main diagonal (i.e. a correlation matrix) with $\frac{1}{2}n(n-1)$ parameters. α and β are scalar parameters, giving a total number of parameters of $0.5n^2 - 0.5n + 2 = 0.5n^2 + O(n) = O(n^2)$. Hence model estimation is feasible only for small to moderately-sized cross-sections, see columns b and e in Table 2.

A less restrictive parameterization for the correlation dynamics is the Generalized DCC in Engle (2002) with dynamics

$$Q_t = R \odot \left(1_n 1_n' - A^{\diamondsuit} - B^{\diamondsuit}\right) + A^{\diamondsuit} \odot u_{t-1}^{\diamondsuit} u_{t-1}^{\diamondsuit\prime} + B^{\diamondsuit} \odot Q_{t-1} \tag{31}$$

where $A^{\diamondsuit}$ and $B^{\diamondsuit}$ are symmetric matrices, each one containing $\frac{1}{2}n(n+1)$ free elements. Both specifications (30) and (31) guarantee Q_t to be p.d. provided $A^{\diamondsuit}$, $B^{\diamondsuit}$ and R are positive (semi-)definite, thanks to properties of Hadamard products, see e.g. Styan (1973), Theorem 3.1. The total number of parameters is hence $n(n+1) + \frac{1}{2}n(n-1) = 1.5n^2 + O(n) = O(n^2)$, i.e. of the same order as (30).

5.10 Structured DCC

A structured DCC specification can be obtained as a restriction to a generalization of (31). Consider the vec of eq. (31), which reads

$$\text{vec}\; Q_t = \text{vec}\; R + A\text{vec}\left(u_{t-1}^{\diamondsuit} u_{t-1}^{\diamondsuit\prime} - R\right) + B\text{vec}\,(Q_{t-1} - R) \tag{32}$$

where $A := \text{diag}(\text{vec } A^{\diamondsuit})$, $B := \text{diag}(\text{vec } B^{\diamondsuit})$, where we have used the property $\text{vec}\,(a \odot b) = \text{diag}(\text{vec } a)\text{vec}\,(b)$. Because of symmetry, just as in the VEC-specification in Subsection 5.3, this can be cast in a vech form as follows:

$$\text{vech } Q_t = \text{vech } R + A^* \text{vech}\left(u_{t-1}^{\diamondsuit} u_{t-1}^{\diamondsuit\prime} - R\right) + B^* \text{vech}\,(Q_{t-1} - R)\,, \tag{33}$$

where A, B and A^*, B^* are restricted as in (22).

The A, B matrices in (32) – or the A^*, B^* matrices in (33) – are assumed to be diagonal; this implies that no conditional correlation spill-over of feedback effects are allowed in the dynamic equation of Q_t. Such effects can be included by de-restricting A, B, A^*, B^* to be possibly non-diagonal; in this case we refer to the DCC model as VEC-DCC.

A structured specification can be obtained as a special case of the VEC-DCC model, by assuming A, B in (32) (respectively A^*, B^* in (33)) to be spatial matrices of the type (8) and $R = \text{cor}(S^{-1}VS^{-1\prime})$, to be a correlation matrix of the type (17)

$$A = A_0 + A_1 W_{n^2}, \qquad B = B_0 + B_1 W_{n^2}, \qquad S = I_n - \Phi W_n, \tag{34}$$

where $A_j := \text{diag}(a^j)$, $B_j := \text{diag}(b^j)$, $j = 0, 1$, $\Phi := \text{diag}(\phi)$, $V := \text{diag}(v)$. The number of unrestricted coefficients in a^j, b^j, ϕ, v are $2n(n+1) + 2n = 2n^2 + O(n) = O(n^2)$, which is quadratic in n and hence unfeasible for estimation on large cross-sections. Group-homogeneous specifications reduce the number of parameters, especially the ones in a^j and b^j; a group-homogeneous specifications for a^j and b^j gives $4k$ parameters, implying a total number of parameters equal to $4k+2n = 2n+O(1) = O(n)$, which is linear in the number of parameters.[12] This specification allows for correlation spill-overs and it is estimable also on large cross-sections, see columns d and g in Table 2.

5.11 Modifications of DCC and structured specifications

Tse and Tsui (2002) introduced a close alternative to DCC, called Varying Conditional Correlation, VCC. As in the CCC and DCC classes, the conditional covariance is decomposed as in (27) and the conditional correlation matrix R_t is assumed to be generated by the dynamics

$$R_t = (1 - \alpha - \beta)\, R + \alpha \Psi_{t-1} + \beta R_{t-1} \tag{35}$$

$$(\Psi_t)_{ij} := \sum_{l=1}^{L} u_{i,t-l}^{\circ} u_{j,t-l}^{\circ}, \qquad u_{i,t-l}^{\circ} := \frac{u_{i,t-l}^{\diamondsuit}}{\sum_{j=1}^{L} u_{i,t-j}^{\diamondsuit}}$$

where Ψ_{t-1} can be considered as a local sample correlation. The model provides positive definite R_t matrices if $L \geq M$. The number of parameters is identical as in the case of (simple, generalized or VEC) DCC. The introduction of structured specifications in the VCC model is identical to the DCC case; it just amounts to substitute R_t for Q_t and Ψ_t for $u_t^{\diamondsuit} u_t^{\diamondsuit\prime}$ in (33) and (32) above. It also gives the same number of parameters in the structured specification.

[12] Here and elsewhere we assume that the number of classes k does not grow with the cross-section dimension n.

Cappiello, Engle and Sheppard (2006) introduced a generalized DCC model where the evolution of Q_t is governed by a BEKK or quadratic type equation

$$Q_t = \left(R - ARA' - BRB'\right) + Au_{t-1}^{\diamondsuit}u_{t-1}^{\diamondsuit\prime}A' + BQ_{t-1}B' \tag{36}$$

where the matrices A and B can be restricted or unrestricted. Again Q_t is positive definite if the constant term is positive definite. For unrestricted A, B, R, the number of parameters is $\frac{3}{2}n(n+1) = 1.5n^2 + O(n) = O(n^2)$, and hence the model cannot be estimated on large cross-sections. Again here one can consider structured A, B, R of the type (29); under these restrictions the number of parameters is $6n = O(n)$, which is linear in n, hence rendering the model estimable also on large cross-sections.

A similar generalization of the VCC model of Tse and Tsui (2002) is

$$R_t = \left(R - ARA' - BRB'\right) + A\Psi_{t-1}A' + BR_{t-1}B'. \tag{37}$$

where an analogous structured specification can be defined. In all the dynamic correlation models the parameter interpretation is similar to the one reported for the structured-BEKK models.

Recently, Engle and Kelly (2008) have introduced a model of dynamic equi-correlation, where $W = J_n$, $R_t = I - \rho_t W$ and ρ_t is taken to be the average of distinct, off-diagonal elements of $\mathrm{cor}(Q_t)$. In light of the above, R_t is seen to be an example of spatial matrix of type (5).

5.12 An example of Likelihood decomposition

We here provide an example of how certain structured specifications imply a decomposition of the joint likelihood into components, each one corresponding to a group of assets. The example is based on the structured GO-GARCH specification of Subsection 5.6.

We consider the group-homogeneous GO-GARCH specification, with $\phi = \left(\phi_1 1'_{n_1} : \cdots : \phi_k 1'_{n_k}\right)'$, where ϕ_i are scalar parameters. We partition $u_t := (u'_{1,t} : \cdots : u'_{k,t})'$ in sub-vectors u_{it} of dimension $n_i \times 1$, each one corresponding to a sector, and partition also $v_t := (v'_{1,t} : \cdots : v'_{k,t})'$, c, $a^j = (\alpha_1^j 1'_{n_1} : \cdots : \alpha_k^j 1'_{n_k})'$, $b^j = (\beta_1^j 1'_{n_1} : \cdots : \beta_k^j 1'_{n_k})'$, $j = 0, 1$ conformably, where α_i^j and β_i^j are scalar parameters. Note that $v_{i,t} := (v_{i,t,1} : \cdots : v_{i,t,n_i})'$ is a $n_i \times 1$ vector. We collect parameters in $\theta_i := (\phi_i : \alpha_i^0 : \alpha_i^1 : \beta_i^0 : \beta_i^1)'$, $\theta := (\theta'_1 : \cdots : \theta'_k)'$ and assume for simplicity that there are no parameters in the conditional mean, i.e. μ_t is known.

Form Subsection 4.3 one has $\log\det S = \sum_{i=1}^k \log\left(\varphi(1, \phi_i, n_i)\right)$; moreover

$$u_t'\Sigma_t^{-1}u_t = \sum_{i=1}^k u_{it}^{*\prime}\mathrm{diag}\left(1_{n_i}./v_{it}\right)u_{it}^*, \qquad u_{it}^* := \left(I_{n_i} - \phi_i J_{n_i}\right)u_{it}.$$

This implies that the log-likelihood function ℓ_t can be decomposed into k blocks, $\ell_t = \sum_{i=1}^k \ell_{it}$, where

$$-2\ell_{it} := \sum_{j=1}^{n_i} \log v_{i,t,j} - 2\log\left(\varphi(1, \phi_i, n_i)\right) + u_{it}^{*\prime}\mathrm{diag}\left(1_{n_i}./v_{it}\right)u_{it}^*.$$

A similar decomposition applies to the dynamics of the variances v_{it}. In fact, one can rewrite (25) as a set of autonomous dynamic equations

$$v_{i,t} = c_i + \left(\alpha_i^0 I_{n_i} + \alpha_i^1 J_{n_i}\right)(u_{i,t-1} \odot u_{i,t-1}) + \left(\beta_i^0 I_{n_i} + \beta_i^1 J_{n_i}\right) v_{i,t-1},$$

where $c_i = (1 - \left(\alpha_i^0 + \beta_i^0\right) - \left(\alpha_i^1 + \beta_i^1\right))1_{n_i}$, $i = 1, \ldots, k$.

When the parameters θ_i and θ_j, $i, j = 1, \ldots, k$, $i \neq j$, are variation independent, likelihood-maximization for the group-homogeneous GO-GARCH specification θ can be decomposed in a set of k optimizations, each one concerning one θ_i and corresponding to a group u_{it}. This property may be of interest in large cross-sections.

5.13 MGARCH with less-straightforward structured specification

Structured specifications are not easy to devise for some MGARCH models, such as for factor-, exponential- and orthogonal- MGARCH.

In fact, factor models do not constrain the type of factor, and the spatial idea as the source of co-movement does not strictly apply. Exponential MGARCH models, Chiu et al. (1996), Kawakatsu (2005), do not have a clear interpretation of parameters in the dynamic equation of conditional covariances. Hence imposing a spatial structure – while straightforward – does not appear to lead to a direct interpretation of parameters, hence falling short of the interpretability criterion. Finally Orthogonal MGARCH models would require to restrict spatial matrices to be orthogonal; while some spatial matrices satisfy this requirement, this is not a general property. We have hence decided not to discuss them here.

6 Other MVM models

In this section we discuss how structured specifications can be applied to other MVM models, outside the class of MGARCH. We discuss Multivariate Stochastic Volatility (MSV) and Multivariate Realized Volatility (MRV) in the next two subsections. Here the likelihood function changes form according to each model, and hence we limit the discussion to the number of parameters.

6.1 Stochastic Volatility

We consider the basic MSV model of Harvey et al. (1994), and refer to Asai, McAleer and Yu (2006) for a more complete review. The model consists of

$$u_t = \Sigma_t^{\frac{1}{2}} \varepsilon_t, \qquad \Sigma_t = X V_t X', \qquad V_t = \exp(\mathrm{diag}(v_t)), \tag{38}$$

$$v_t = \mu + b \odot v_{t-1} + \eta_t, \tag{39}$$

where b, μ, v_t, ε_t and η_t are $n \times 1$ vectors and X is an $n \times n$ matrix of full rank. The stochastic vectors ε_t and η_t are independent and jointly normal, they have 0 expectations and $\mathbb{V}(\varepsilon_t) = R_\varepsilon$, $\mathbb{V}(\eta_t) = \Sigma_\eta$; R_ε is a correlation matrix.

The parameter are X, b, μ, R_ε and Σ_η; if no restriction applies to these parameters, the total count is $n^2 + 2n + \frac{1}{2}n(n-1) + \frac{1}{2}n(n+1) = 2n^2 + O(n) = O(n^2)$, which grows as n^2. This contributes to make these models non-estimable on medium-to-large cross-sections.

We consider a generalization of (39) that allows for variance spill-over; more precisely

$$v_t = \mu + Bv_{t-1} + \eta_t, \tag{40}$$

where B is an $n \times n$ matrix. When B is unrestricted, the parameter count grows, but remains of the same $O(n^2)$ order of magnitude.

Structured specification can be obtained assuming Σ_t in (38) to be an instance of a SAR covariance structure $\Sigma_t = S^{-1}V_tS^{-1\prime}$, see eq. (4), i.e. $X = S^{-1}$ in (38), where $S = I - \Phi W_n$, $\Phi = \mathrm{diag}(\phi)$. The correlation matrix R_ε and the covariance matrix Σ_η can also be modelled as resulting from SAR processes, setting

$$\Sigma_\eta = S_\eta^{-1}V_\eta S_\eta^{-1\prime}, \qquad R_\varepsilon = \mathrm{cor}(S_\varepsilon^{-1}V_\varepsilon S_\varepsilon^{-1\prime})$$

where $S_j = I - \mathrm{diag}(\phi_j)W_n$, $V_j = \mathrm{diag}(v_j)$, $j = \eta, \varepsilon$. Finally the B matrix can be taken to be a spatial matrix of the type $B_0 + B_1W_n$, with $B_j = \mathrm{diag}(b^j)$, $j = 0, 1$.

The parameters of the structured-MSV specification are ϕ_j, v_j, for $j = \eta, \varepsilon$, ϕ, μ and b^j for $j = 0, 1$, which, if left unrestricted, give a total number of parameters equal to $8n = O(n)$ which is linear in n. This makes the structured-MSV specification potentially more amenable to estimation also on large cross-sections. Again the interpretation of the parameters in the spatial matrices is analogous to the case described for the structured-BEKK specification.

6.2 Models for realized volatility

Multivariate Realized Volatility models (MRV) describe the dynamics of realized volatility.[13] As a representative of this class, we consider the Wishart-based model proposed by Gourieroux (2006), Gourieroux, Jasiak and Sufana (2008). In this model class, one assumes to observe a p.d. $n \times n$ process Y_t, measurable with respect to the filtration $\mathcal{F}_t$. Conditionally on $\mathcal{F}_{t-1}$, Y_t is assumed to be noncentral Wishart with m degrees of freedom and conditional expectation equal to

$$\mathbb{E}_{t-1}(Y_t) = AY_{t-1}A' + m\Sigma \tag{41}$$

A is interpreted as an AR parameter matrix, Σ is a p.d. matrix and $m > n - 1$ is a real parameter, see Gourieroux, Jasiak and Sufana (2008). The number of parameters in (41), if A and Σ are unrestricted, is $n^2 + \frac{1}{2}n(n+1) + 1 = 1.5n^2 + O(n) = O(n^2)$.

Structured specifications can be introduced assuming that the parameter matrix A belongs to the class of spatial matrices; moreover, Σ can be assumed to be the covariance matrix of a SAR model, i.e.

$$\Sigma = S^{-1}VS^{-1\prime}, \qquad S = I_n - \Phi W_n, \qquad A = A_0 + A_1W_n,$$

[13] Realized Volatility (RV) is also an estimator of some underlying continuous time process, see McAleer and Medeiros (2008) for a review. Here we are interested in the direct modeling of RV.

where $A_j := \text{diag}(a^j)$, $j = 0, 1$, $\Phi = \text{diag}(\phi)$, $V = \text{diag}(v)$. The number of unrestricted coefficients in a^0, a^1, ϕ, v is $4n = O(n)$, which is linear in n and hence feasible for estimation also on large cross-sections.

The Wishart model is similar to a BEKK model without a ARCH term; as a result, parameter interpretation is close to that reported for the structured-BEKK specification. Wishart densities have also been used in Factor models, see for instance Philipov and Glickman (2006); see also Bauer and Vorkink (2007), Chiriac and Voev (2008).

Similar structured specification can be devised for the dynamics of Vector Multiplicative Error Models (MEM) of Cipollini, Engle and Gallo (2006).

7 Extensions

The previous sections show the importance of weight matrices W in structured MVM specifications. The first observation is that weight matrices W only need to be measurable with respect to $\mathcal{F}_{t-1}$, the information set available at time $t-1$. Hence weight matrices W can be both time-varying and stochastic.

In the rest of this section we extend the discussion on how to define weight matrices, using results in graph theory (Subsection 7.1), exploiting covariates (Subsection 7.2) or requiring invariance (Subsection 7.3). We present results only for one generic weight matrix W_t, using the example of n assets.

7.1 Graph theory

Weight matrices can be obtained as a by-product of the definition of a simple graph on the assets; see e.g. Bondy and Murty (1976) for an introduction to graphs. Each of the n assets represents a vertex of a graph; edges between pairs of vertices represent connections among assets.[14] Edges may be directed or undirected; if undirected, the adjacency matrix associated with the graph is symmetric.

The construction of edges for a single classification criterion of assets (like industrial sectors) corresponds to undirected edges for asset i pointing to assets j from the same sector. Note that the concept of directed edges allows more flexibility in this context, and allows to have some influence from i to j but not vice-versa. The construction of edges may reflect a number of other characteristics in addition to the fact that i and j correspond to assets in the same sector.

Once a (directed or undirected) graph has been defined for a given t (consisting of vertices and edges), one can consider the adjacency matrix A with element a_{ij} equal to 1 if there is an edge from i to j and 0 otherwise. A weight matrix can then be constructed by setting $W_t^* := A'$ and row-normalizing W_t^* into W_t. The transposition is needed here in order to respect the source-destination convention in adjacency matrices; for instance, row 1 of A'

[14] When only one edge connects two vertices the graph is called simple.

contains a 1 in column j for which there is a vertex from j to 1. Obviously, the transposition is immaterial for undirected graphs.

Graph theory may hence be used to construct weight matrices in a more general form that the one presented in Section 3 for classification criteria, which are a special case. For more discussion of the links between graphs and spatial models, see Martellosio (2008) and reference therein.

7.2 Covariates

In this subsection we discuss how one can convert information on covariates x_{t-1} (contained in $\mathcal{F}_{t-1}$) into an aggregate proximity weight matrix W_t. This allows to associate W_t to firm-specific time-varying covariates such as market value, book value, momentum, earnings/price, cash-flow/price, dividend yield, short- and long-term reversals.

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 $x := q^{-\frac{1}{b}} \left\| x_{i,t-1} - x_{j,t-1} \right\|_b$, where $\|a\|_b := \left(\sum_{i=1}^{q} |a_i|^b \right)^{1/b}$; take here for simplicity $b = 1$. Next define the un-normalized weight matrix $W_t^* := (w_{i,j,t}^*)$ with weights

$$w_{i,j,t}^* := (1 - \delta_{i,j})\, f(sx), \qquad f(y) := \exp\left(-y^r\right), \qquad r > 0, \tag{42}$$

where $\delta_{i,j}$ is Kronecker's index, which takes value 1 for $i = j$, and r, s are positive constants. For instance $r = 16$, $s = 10$. The normalized weight matrix $W_t := (w_{i,j,t})$ is obtained by row-normalization of W^*, $w_{i,j,t} := w_{i,j,t}^* / \left(\sum_{j=1}^{n} w_{i,j,t}^* \right)$.

The choice of values for r and s is linked to the following interpretation of (42):

1. x is an average of the distance between $x_{i,t-1}$ and $x_{j,t-1}$;

2. $y := sx$ re-scales the aggregate indicator on a $[0, s]$ scale; for $s = 10$ one has one has $0 \leq y \leq 1$ when $0 \leq x \leq \frac{1}{10}$; changing s dilates the scale; $s = 20$ gives values of $0 \leq y \leq 1$ for $0 \leq x \leq \frac{1}{20}$;

3. the function $\exp(-y^r)$ maps $[0, s]$ to $(0, 1]$ in a monotonic decreasing way; values of $0 < y \leq 1$ are mapped into values close to 1 and values $1 < y$ are mapped into values close to 0; the higher the exponent r, the more this map smoothly approximates $1(0 \leq y \leq 1)$, where $1(\cdot)$ is the indicator function.

Note that in step 1, b could be chosen > 1 or that x could be replaced by any other mean value, like the median, or by other weighted averages. The choice of s determines the 'threshold' of the indicator function in point 3; $s = 10$ sets the threshold at the 0.1 point in the x scale, $s = 20$ sets it at 0.05 etc. For any given value of s, higher values of r deliver a function $\exp(-y^r)$ that resembles more closely $1(0 \leq y \leq 1)$. Small values of r instead give a function $\exp(-y^r)$ with steep decrease in the proximity of 0. Hence the choice of r and s establishes with range of x values gives weights $w_{i,j,t}^*$ close to 1.

We note that this definition of 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 (42) provides a generalized on how to define weight matrices.

The choice $f(y) := \exp(-y^r)$ in (42) 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 and products like Wu should be computed including all elements, unlike when W is sparse. A different choice that prevents this phenomenon is to replace $\exp(-y^r)$ with $\exp(-y^r)\,1(y < c)$ say, that sets all weights equal to 0 for $y = sx \geq c$. When for instance $s = 8$, one can set $c = 2$, because $\exp\left(-2^8\right)$ is of the order 1.5×10^{111}, a very small number.

Values of r and s in (42) can be fixed in advance; alternatively, they could be included as parameters to be estimated. When using the model in prediction, one could assume that W_t is fixed at the last available value, or that it changes according to scenarios designed to reflect possible evolutions in the market. Note that in principle, the number of indicators q could also change over time.

This discussion shows that any discrete or continuous, time-varying or time-constant covariate can be used to define weight matrices.

7.3 Invariance

Invariance is often used to construct parsimonious models in multivariate analysis; see for instance Andersson (1975) for the theory of invariant covariance models for the multivariate normal distribution. Invariance arguments can be used to restrict the set of possible MVM structures, which includes the definition of the weight matrix W_t. In this subsection we illustrate this through an example with a single classification criterion F associated to assets' sectors. We consider a homogeneous structured GO-GARCH specification, see Subsection 5.6.

Assume that one wishes to impose invariance of the process with respect to the transformation $u_t \curvearrowright Pu_t$, where P is any matrix that permutes columns i and j, where i and j correspond to assets from the same sector; we call this situation P-invariance. For instance, for assets 1 and 2, $P = (e_2 : e_1 : e_3 : \cdots : e_n)$, where e_i is the i-th column of the identity matrix. Note that P is symmetric orthogonal, $P = P' = P^{-1}$.

P-invariance implies that $\Sigma_t = P\Sigma_t P$, where $\Sigma_t = (I - \phi W_t)^{-1} V_t (I - \phi W_t)^{-1\prime}$. One sufficient condition for P-invariance is to assume that W_t and V_t are satisfy $W_t = PW_tP$ and $V_t = PV_tP$, because

$$P\Sigma_t P = (I - \phi PW_tP)^{-1} PV_tP (I - \phi PW_tP)^{-1\prime} = (I - \phi W_t)^{-1} V_t (I - \phi W_t)^{-1\prime} = \Sigma_t.$$

The conditions $W_t = PW_tP$ and $V_t = PV_tP$ imply that $w_{ij,t} = w_{ji,t}$ and $v_{i,t} = v_{j,t}$, when i and j are from the same sector. The requirement $w_{ij,t} = w_{ji,t}$ implies that the weight matrix W_t is of the form (10). Recall also that $v_t = c + A(u_{t-1} \odot u_{t-1}) + Bv_{t-1}$, and hence $Pv_t = Pc + PAP(P(u_{t-1} \odot u_{t-1})) + PBPv_{t-1}$. This implies that $Pv_t = v_t$ if and only if

$$Pc = c, \qquad PAP = A, \qquad PBP = B \tag{43}$$

For a homogeneous structured specification one has $A = a^0 I + a^1 W_t$, $B = b^0 I + b^1 W_t$, for scalar a^j, b^j, $j = 0, 1$; one finds

$$PAP = a^0 I + a^1 P W_t P = a^0 I + a^1 W_t = A$$

by the invariance of the weight matrix $PW_tP = W_t$. Similarly for $B = b^0 I + b^1 W_t$ one has $PBP = B$. Hence the only additional restriction in (43) is that $Pc = c$, i.e. the intercept in the GARCH equation is identical for stocks in the same sector.

This example shows that, imposing a certain type of invariance, one obtains a specific type of weight matrix. Different choices of P give rise to different weight matrices.

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. Structured specifications form an interesting class for volatility models, because they provide both flexible and parsimonious models, allowing for variance spill-over and being characterized by a number of parameters that grows linearly with the cross-section dimension. Moreover structured specifications parameters have a direct economic interpretation that reflects the chosen notion of economic proximity. Finally, structured specifications can help in speeding up computations for model estimation.

References

Aielli G. P. (2008), Consistent estimation of large scale dynamic conditional correlations, Quaderno di economia, statistica e analisi del territorio, Università di Messina, Facoltà di Scienze Politiche, n 47.

Aguilar, O. and M. West (2000), Bayesian dynamic factor models and portfolio allocation. Journal of Business and Economic Statistics 18:338–357.

Alexander, C.O. (2001), Market Models. New York: Wiley.

Alexander, C.O. and A.M. Chibumba (1997), Multivariate orthogonal factor GARCH. University of Sussex, Mimeo

Andersson S. (1975), Invariant Normal Models, The Annals of Statistics 3, 132-154.

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

Anselin L. (2002), Under the hood. Issues in the specification and the interpretation of spatial regression models, Agricultural Economics 27, 247-267.

Asai, M., M. McAleer and J. Yu (2006), Multivariate Stochastic Volatility: a review, Econometric Reviews, 25-(2-3), 145-175

Asai, M., M. Caporin and M. McAleer (2008), Block structure multivariate stochastic volatility models, unpublished manuscript

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

Baltagi B.H., Bresson G., and Pirotte A. (2007), Panel Unit Root Tests and Spatial Dependence, Journal of Applied Econometrics, 22, 339-360.

Bauer, G. and K. Vorkink (2007), Multivariate realized stock market volatility, Bank of Canada Working Paper 2007-20.

Bera A. K. and P. Simlai (2004), Testing for Spatial Dependence and a Formulation of Spatial ARCH (SARCH) Model with Applications, working paper University of Illinois.

Billio, M., M. Caporin and M. Gobbo (2006), Flexible dynamic conditional correlation multivariate GARCH models for asset allocation, Applied Financial Economic Letters, 2-2, 123-130

Billio, M. and M. Caporin, (2009), A generalized dynamic conditional correlation model for portfolio risk evaluation, Mathematics and Computers in Simulation, forthcoming

Bollerslev, T. (1990), Modelling the coherence in short-run nominal exchange rates: a multivariate generalized ARCH approach. Review of Economics and Statistics 72:498–505.

Bollerslev, T., R. Engle, and J. Wooldridge (1988), A Capital Asset Pricing Model with Time Varying Covariances, Journal of Political Economy, 96, 116–131.

Bollerslev, T. (2008), Glossary to ARCH (GARCH), available at http://www.princeton.edu/~mwatson/ef/glossary_arch.pdf

Bonato, M., M. Caporin and A. Ranaldo (2008), Forecasting realized (co)variances with Block Wishart Autoregressive Models, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1282254

Bondy J.A. and U.S.R. Murty (1976) Graph Theory with Applications, Macmillan Press, available at http://www.ecp6.jussieu.fr/pageperso/bondy/books/gtwa/gtwa.html

Caporin, M. and M. McAleer (2008), Scalar BEKK and Indirect DCC, Journal of Forecasting, 27-6, 537-549

Cappiello, L., R. Engle and K. Sheppard (2006), Asymmetric dynamics in the correlations of global equity and bond returns, Journal of Financial Econometrics, 4-4, 537-572.

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

Chiriac, R. and V. Voev (2008), Modelling and forecasting multivariate realized volatility, CREATES Research paper 2008-39

Cipollini F., R.F. Engle, G.M. Gallo (2006), Vector multiplicative error models: representation and inference, NBER WP 12690.

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

Conley T.G. and Ligon E. (2002) Economic Distance and Cross-Country Spillovers. Journal of Economic Growth 7, 157–187.

Conley T.G. and Topa G (2002), Socio-economic distance and spatial patterns in unemployment. Journal of Applied Econometrics: 17: 303–327. DOI: 10.1002/jae.670.

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

Chiu, T. Y. M., T. Leonard and K.-W. Tsui (1996), The matrix-logarithmic covariance model. Journal of the American Statistical Association 91:198–210.

Diebold, F. X., and M. Nerlove (1989), The dynamics of exchange rate volatility: a multivariate latent factor ARCH model. Journal of Applied Econometrics 4:1–22.

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

Engle, R. F. (2002), Dynamic conditional correlation: a simple class of multivariate generalized autoregressive conditional heteroskedasticity models. Journal of Business and Economic Statistics 20:339–350.

Engle, R. F. and B. Kerry (2008), Dynamic equicorrelation, working paper, available at http://pages.stern.nyu.edu/~rengle/Dynamic%20Equicorrelation.pdf

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

Engle, R. F., V. K. Ng, and M. Rothschild (1990), Asset Pricing with a Factor-ARCH Covariance Structure: Empirical Estimates for Treasury Bills, Journal of Econometrics, 45, 213–238.

Fernandes, M., B. de Sá Mota and G. Rocha (2005), A mulivariate conditional autoregressive range model, Economics Letters, 86, 435-440

Gall, J., G. Pap and M. van Zuijlen (2004), Maximum likelihood estimator of the volatility of forward rates driven by geometric spatial AR sheet. Journal of Applied Mathematics 4, 293-309.

Gourieroux, C. (2006), Continuous time Wishart process for stochastic risk, Econometric Reviews 25(2–3):177–217.

Gourieroux, C., J. Jasiak and R. Sufana (2008), The Wishart autoregressive process of multivariate stochastic volatility, Journal of Econometrics, forthcoming

Harvey, A. C., E. Ruiz and N. Shephard (1994), Multivariate stochastic variance models. Review of Economic Studies 61:247–264.

Huse, C. (2006), Term structure modelling with spatial dependence and observable state variables, paper presented at the International Workshop on Spatial Econometrics and Statistics, Rome.

Kawakatsu, H. (2005), Matrix exponential GARCH. Journal of Econometrics 134, 95–128.

Jacquier, E., N.G. Polson and P.E. Rossi (1995), Models and priors for multivariate stochastic volatility. CIRANO Working Paper No. 95s-18, Montreal.

Jacquier, E., N.G. Polson and P.E. Rossi (1999), Stochastic volatility: univariate and multivariate extensions. CIRANO Working paper 99s-26, Montreal.

Lee L.F. (2002), Consistency and efficiency of least squares estimation for mixed regressive, spatial autoregressive models, Econometric theory 18,252-277.

Lee L.F. (2004), Asymptotic distribution of Quasi-Maximum Likelihood for spatial autoregressive models. Econometrica 72, 1899-1925.

Lin, W. (1992), Alternative Estimators for Factor GARCH Models - A Monte Carlo Comparison, Journal of Applied Econometrics, 7, 259–279.

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

Linton O. (2008) Semiparametric and Nonparametric ARCH Modelling, in: T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch (eds.), Handbook of Financial Times Series. New York: Springer.

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

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

Magnus J.R. and C. Müris (2008), specification of variance matrices for panel data models, working paper available at http://center.uvt.nl/staff/magnus/

Magnus, J.R. and H. Neudecker (1988), Matrix differential calculus with applications in statistics and econometrics, Chichester, John Wiley & Sons

Marcus, M. and N.A. Khan (1959), A note on the Hadamard product, Canadian Mathematical Bulletin, 2, 81-83

Martellosio F. (2008), Power properties of invariant tests for spatial autocorrelation in linear regression, forthcoming in Econometric Theory, available at http://mpra.ub.uni-muenchen.de/7255/

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), Realized volatility: a review, Econometric Reviews, 27-(1-3), 10-45.

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

Pelletier, D. (2006), Regime switching for dynamic correlations, Journal of Econometrics, 131-(1-2), 445-473

Pesaran M. H., T. Schuermann and Weiner S. (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 (2008), Large panels with common factors and spatial correlations, available at http://www.econ.cam.ac.uk/dae/repec/cam/pdf/cwpe0743.pdf

Philipov, A., and M.E. Glickman (2004), Multivariate stochastic volatility via Wishart processes. To appear in Journal of Business and Economic Statistics.

Philipov, A., and M.E. Glickman (2006), Factor multivariate stochastic volatility via Wishart processes. Econometric Reviews 25(2–3):311–334.

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

Silvennoinen A. and T. Teräsvirta (2007), Modelling Multivariate Autoregressive Conditional Heteroskedasticity with the Double Smooth Transition Conditional Correlation GARCH Model, available at http://swopec.hhs.se/hastef/papers/hastef0652.pdf

Silvennoinen A. and T. Teräsvirta (2008), Multivariate GARCH models, in: T.G. Andersen, R.A. Davis, J.-P. Kreiss and T. Mikosch (eds.), Handbook of Financial Times Series. New York: Springer.

Styan, G.P.H. (1973), Hadamard products and multivariate statistical analysis, Linear Algebra and its Applications, 6, 217-240

Van der Weide, R. (2002), GO-GARCH: A Multivariate Generalized Orthogonal GARCH Model, Journal of Applied Econometrics, 17, 549–564.

Vrontos I.D., P. Dellaportas and D.N. Politis (2003), A full-factor multivariate GARCH model. Econometrics Journal 6, 311–333.

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

Yan J. (2006) Spatial stochastic volatility for lattice data, working paper University of Iowa.

A Appendix

A.1 Spatial and weight matrices

Here we provide definitions of weight matrices and spatial matrices, along with some of their properties. Note that $W_{ij} = 0$ need not imply $W_{ji} = 0$, i.e. that neighborhood relations need not be symmetric. We concentrate attention to row-normalized weight matrices W, which are defined as a subset of the set of stochastic matrices.

The normalization of W to have row sums equal to 1 is not restrictive in the class of heterogeneous SAR processes in (6) below. In fact let W be un-normalized and observe that $W^* := \text{diag}(1./W1_n)W$ is the corresponding normalized weight matrix; next note that $\text{diag}(\phi)W$ in (6) can be rewritten as $\text{diag}(\phi^*)W^*$ with $\phi^* = \phi \odot W1_n$. Hence one can assume W to be normalized for heterogeneous SAR processes.

The heterogeneous specification also allows to accommodate the case where one unit has no neighbors, i.e. a subset of the rows in $\text{diag}(\phi)W$ is equal to 0. This can be accomplished either by setting the corresponding rows in W to be zero or by setting the corresponding element in ϕ to 0, and setting one non-diagonal element equal to 1 in each corresponding row of W. We prefer the latter option, which allows to maintain the hypothesis that all rows of W have some nonzero element.

For homogenous SAR processes, the row-normalization of W can be restrictive. In case all row sums are equal to a constant c (a situation which excludes the possibility of a subset of rows to be equal to 0) then ϕW can be rewritten as $\phi^* W^*$ with $\phi^* = \phi c$ and $W^* := c^{-1}W$. In this case the normalization can be adsorbed into the AR coefficient ϕ^*. There are however situations where one may be interested in non-normalized W matrices, see Martellosio (2008).

In this appendix we use the following notation: $\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{R}_{0+}$ are the set of all, positive and nonnegative real numbers; $\mathbb{A}^{n\times m}$ indicates the set of all matrices of dimensions $n \times m$ whose entries belong to the set $\mathbb{A}$; $\mathbb{A}^n := \mathbb{A}^{n\times 1}$; 1_n indicates the $n \times 1$ vector with all entries equal to 1, I_n indicates the identity matrix of order n. We use the vec and the Kronecker product $\otimes$ operators as in Magnus and Neudecker (1988) and define $\text{diag}(x)$, where x is $n \times 1$, as the square diagonal matrix with x on the main diagonal. Element i, j of matrix A is indicated as $(A)_{ij}$.

Definition 2 (Weight matrices) *The class of stochastic matrices $\mathcal{P}_n$ is defined as the class of square matrices of order n with nonnegative real entries, such that all rows sum to 1, $\mathcal{P}_n := \{P \in \mathbb{R}_{0+}^{n\times n} : P1_n = 1_n\}$. The class of weight matrices $\mathcal{W}_n$ is the subset of $\mathcal{P}_n$, $\mathcal{W}_n \subset \mathcal{P}_n$, of matrices with 0 diagonal entries, $\mathcal{W}_n := \{W \in \mathcal{P}_n : (W)_{ii} = 0, i = 1, \ldots, n\}$.*

An example of weight matrix is the matrix $J_n := (n-1)^{-1}(1_n 1_n' - I_n) \in \mathcal{W}_n$. The definition of CAR processes requires 0 entries on the diagonal of W; this restriction also makes sense in SAR processes, see Martellosio (2008) Appendix A.[15]

[15] Martellosio (2008) also assumes that W is an irreducible matrix, which corresponds to assuming that any

A.2 Properties

The class $\mathcal{W}_n$ is a subset of the stochastic matrices $\mathcal{P}_n$, from which it inherits several simple properties. Some of these properties are listed in the following 2 propositions.

Proposition 3 (Convex combinations and products) *Let $\mathcal{A}_n$ indicate either $\mathcal{W}_n$ or $\mathcal{P}_n$, and let $W_i \in \mathcal{A}_n$, $i = 1, \dots, k$. Then the following properties hold:*

1. *$\sum_{i=1}^{k} c_i W_i \in \mathcal{A}_n$ for $0 \leq c_i \leq 1$, $\sum_{i=1}^{k} c_i = 1$ both for $\mathcal{A}_n = \mathcal{W}_n, \mathcal{P}_n$ ($\mathcal{W}_n$ and $\mathcal{P}_n$ are closed under convex combinations);*

2. *$\prod_{i=1}^{k} W_i \in \mathcal{A}_n$ only for $\mathcal{A}_n = \mathcal{P}_n$ (only $\mathcal{P}_n$ is closed under matrix multiplication).*

We next analyze properties of Kronecker products of stochastic matrices in the following Proposition 4. 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 suggests that $P_1 \otimes P_2 \in \mathcal{P}_{n^2}$ and the next proposition discusses to what extent this is true also for $\mathcal{W}_{n^2} \subset \mathcal{P}_{n^2}$.

Proposition 4 (Kronecker products) *Let $\mathcal{A}_n$ indicate either $\mathcal{W}_n$ or $\mathcal{P}_n$, and let $W_1, W_2 \in \mathcal{A}_n$. Moreover let H, K denote generic $n \times n$ matrices. Then the following holds:*

1. *the Kronecker product of two matrices in $\mathcal{W}_n$ and $\mathcal{P}_n$ generates elements of the same class but with dimension n^2, i.e. $W_1 \otimes W_2 \in \mathcal{A}_{n^2}$ both for $\mathcal{A}_{n^2} = \mathcal{W}_{n^2}, \mathcal{P}_{n^2}$;*

2. *moreover $W_1 \otimes W_2 \in \mathcal{W}_{n^2}$ for $W_1 \in \mathcal{W}_n$, $W_2 \in \mathcal{P}_n$ or $W_1 \in \mathcal{P}_n$, $W_2 \in \mathcal{W}_n$.*

3. *Conversely, $H \otimes K \in \mathcal{P}_{n^2}$ implies $H = cW_1$ and $K = \frac{1}{c} W_2$, $c \in \mathbb{R} \backslash \{0\}$, and $W_1, W_2 \in \mathcal{P}_n$ (i.e. if a matrix with Kronecker-product structure is stochastic, then its two matrix factors are proportional to stochastic matrices, with reciprocal constants of proportionality);*

4. *moreover $H \otimes K \in \mathcal{W}_{n^2}$ implies $H = cW_1$ and $K = \frac{1}{c} W_2$ and either $W_1 \in \mathcal{P}_n$, $W_2 \in \mathcal{W}_n$ or $W_1 \in \mathcal{W}_n$, $W_2 \in \mathcal{P}_n$ (i.e. if a matrix with Kronecker-product structure is a weight matrix, then in addition to 3., one of the two matrix factors has zeros on the main diagonal).*

In the following we concentrate attention on the class $\mathcal{W}_n$. We next define the class of spatial matrices $\mathcal{S}_n$ as (a generalization of) the class containing linear combinations of the identity and weight matrices $\{W_i\}_{i=1}^{k}$ with $W_i \in \mathcal{W}_n$. We indicate a given set of weight matrices by $\mathbb{W} := \{W_i\}_{i=1}^{k}$.

vertex in the graph with adjacency matrix corresponding to W has a path to any other vertex.

Definition 5 (Spatial matrices) *The class of spatial matrices* $\mathcal{S}_n := \mathcal{S}_n(\mathbb{W})$ *with respect to a given set of weight matrices* $\mathbb{W}:=\{W_i\}_{i=1}^k$ *with* $W_i \in \mathcal{W}_n, W_0 := I_n$ *is defined as follows:*

$$\mathcal{S}_n := \{A \in \mathbb{R}^{n\times n} : A = \sum_{i=0}^{k} A_i W_i, A_i = \mathrm{diag}(a_i), a_i \in \mathbb{R}^n\}.$$

One can define the subset of $\mathcal{S}_n$ *corresponding to vectors* a_i *with identical entries,* $a_i = \alpha_i 1_n$ *with* $\alpha_i \in \mathbb{R}$*, called the 'homogeneous' spatial matrix class, indicated as* $\mathcal{S}_n^H$, $\mathcal{S}_n^H \subset \mathcal{S}_n$.

We note that if $A \in \mathcal{S}_n$ then $I - A \in \mathcal{S}_n$ and vice versa. One has

Proposition 6 *Let* $\mathbb{I} := \{0, 1, \ldots, k\}$*; if* $A, B \in \mathcal{S}_n(\mathbb{W})$*, then*

$$A \otimes B \in \mathcal{S}_{n^2}(\mathbb{W}^*)$$

where $\mathbb{W}^* := \{W_h^*\}_{h=1}^m$, $W_h^* \in \mathcal{W}_{n^2}$ *with* $W_h^* := W_i \otimes W_j$, $(i,j) = \mathbb{I}^2\backslash\{(0,0)\}$, $h := (k+1)i + j$, $m := (k+1)^2 - 1$.

This proposition is the key in proving that a structured BEKK is a special case of a structured VEC specification.

A.3 Proofs

Proof. Proposition 1. Element w_{hv}^* in (14) 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 $(W_n \otimes W_n)_{hv} = w_{il}^{\diamondsuit} w_{jm}^{\diamondsuit}$ where $w_{ij}^{\diamondsuit} := (W)_{ij}$; hence, using (11), one sees that $w_{hv} = w_{il}^{\diamondsuit} w_{jm}^{\diamondsuit}$. ■

Proof. of Proposition 3. The first two statements are well known for stochastic matrices, see e.g. Seneta (1981, Chapter 4). Here we report proofs for both $\mathcal{A}_n = \mathcal{W}_n, \mathcal{P}_n$ for completeness.

1. $\sum_{i=1}^k c_i W_i \in \mathbb{R}_{0+}^{n\times n}$ and $\sum_{i=1}^k c_i W_i 1_n = \sum_{i=1}^k c_i 1_n = \left(\sum_{i=1}^k c_i\right) 1_n = 1_n$ where the first equality holds by $W_i \in \mathcal{W}_n^*$, and the last one by $\sum_{i=1}^k c_i = 1$. Hence $W_i \in \mathcal{P}_n \forall i$ implies $\sum_{i=1}^k c_i W_i \in \mathcal{P}_n$. If $W_i \in \mathcal{W}_n \forall i$, i.e. all W_i have 0 elements on the diagonal, so does $\sum_{i=1}^k c_i W_i$, so that $W_i \in \mathcal{W}_n \forall i$ implies $\sum_{i=1}^k c_i W_i \in \mathcal{W}_n$.

2. $H_k := \prod_{i=1}^{k} W_i \in \mathbb{R}_{0+}^{n\times n}$. Note that $H_k = H_{k-1}W_k$ and that $H_1 = W_1 \in \mathcal{P}_n$ by hypothesis. Next proceed by induction and assume that $H_i \in \mathcal{P}_n$, and find $H_{i+1}1_n = H_iW_{i+1}1_n = H_i1_n = 1_n$. This shows that $W_i \in \mathcal{P}_n \forall i$ implies $\prod_{i=1}^{k} W_i \in \mathcal{P}_n$. Note that if $W_i \in \mathcal{W}_n \forall i$, i.e. all W_i have 0 elements on the diagonal, this does not imply that $\prod_{i=1}^{k} W_i \in \mathcal{W}_n$, so that this property does not hold for $\mathcal{W}_n$.

■

Proof. of Proposition 4

1. $W_1 \otimes W_2 \in \mathbb{R}_{0+}^{n^2\times n^2}$. Moreover $(W_1 \otimes W_2)\, 1_{n^2} = \mathrm{vec}\,(W_2 1_n 1_n' W_1') = \mathrm{vec}\,(1_n 1_n') = 1_{n^2}$ and hence $W_1 \otimes W_2 \in \mathcal{P}_{n^2}$.

2. If W_1 or $W_2 \in \mathcal{W}_n$, i.e. they have 0 elements on the main diagonal, so does $W_1 \otimes W_2$; hence W_1 or $W_2 \in \mathcal{W}_n$ implies $W_1 \otimes W_2 \in \mathcal{W}_n$.

3. If $H \otimes K \in \mathcal{P}_{n^2}$, then $(H \otimes K)\, 1_{n^2} = \mathrm{vec}\,(W_2 1_n 1_n' W_1') = \mathrm{vec}\,(1_n 1_n') = 1_{n^2}$ and hence

$$H 1_n 1_n' K' = 1_n 1_n', \tag{44}$$

which implies $H1_n, K1_n \in \mathrm{col}(1_n)$, i.e. $H1_n = 1_n c_H$, $K1_n = 1_n c_K$. This shows that $\frac{1}{c_H}H =: W_1$ and $\frac{1}{c_K}K =: W_2$ are stochastic matrices, $W_1, W_2 \in \mathcal{P}_n$. Substituting back in (44) one also finds $c_H c_K = 1$.

4. If $H \otimes K \in \mathcal{W}_{n^2}$, then $H \otimes K \in \mathcal{P}_{n^2}$ and 3. applies. Moreover, the elements on the main diagonal of $\mathrm{dg}\,(H \otimes K) = 0$, which implies $\mathrm{dg}\,(H) \otimes \mathrm{dg}\,(K) = 0$; the latter can hold only if either $\mathrm{dg}\,(K)$ or $\mathrm{dg}\,(K)$ equals 0, i.e. W_1 or $W_2 \in \mathcal{W}_n$.

■

Proof. of proposition 6. 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 $\mathbb{W}:=\{W_i\}_{i=1}^{k}$. One has

$$A \otimes B = \sum_{i,j=0}^{k} A_i W_i \otimes B_j W_j = \sum_{i,j=0}^{k} (A_i \otimes B_j)\,(W_i \otimes W_j) = \sum_{h=0}^{m} C_h W_h^* \tag{45}$$

where $C_h := A_i \otimes B_j$ and $W_h^* := W_i \otimes W_j$ for $h = (k+1)i + j$. By Proposition 3.2 one has $W_h^* \in \mathcal{W}_{n^2}$ unless $h = 0$, for which $W_0^* = W_0 \otimes W_0 = I_n \otimes I_n = I_{n^2}$. Hence $\mathbb{W}^* := \{W_h^*\}_{h=1}^{m}$ is a set of weight matrices in $\mathcal{W}_{n^2}$. This proves the statement by 3.1 if $A_i = \alpha_i I_n$ and $B_i = \beta_i I_n$ are scalar matrices.

In order to prove the statement for generic A_i, B_i, let $A_i =: \mathrm{diag}(a_i)$, $B_i =: \mathrm{diag}(b_i)$ and denote $c_h := a_i \otimes b_j \in \mathbb{R}^{n^2}$. Note that

$$C_h := A_i \otimes B_j = \mathrm{diag}(a_i) \otimes \mathrm{diag}(b_j) = \mathrm{diag}(a_i \otimes b_j) =: \mathrm{diag}(c_h)$$

and hence (45) is a representation conformable with Definition 5. This completes the proof.

■