Matrix Inequality Constraints for Vector (Asymmetric Power) GARCH Models and MEM with spillovers: some New (Mixture) Formulations

M. Karanasos[†*], Y. Xu[‡]
[†]Brunel University London, UK;[‡] Cardiff University

November 28, 2017

Abstract

In this paper we review and generalize results on the derivation of tractable non-negativity (necessary and sufficient) conditions for N-dimensional asymmetric power GARCH/HEAVY models and MEM. We show that these non-negativity constraints are translated into simple matrix inequalities, which are easily handled. One main concern is that the existence of such conditions is often ignored by researchers. We hope that our paper will create more awareness of the presence of these non-negativity conditions and increase their usage. In practice these constraints may not be fulfilled. To handle these cases we propose a new mixture formulation in order to eliminate some of these constraints. By using the exponential specification for some (but not all) of the conditional variables in the system we considerably reduce the dimensions of them. We also obtain new theoretical results about the second moment structure and the optimal forecasts of such multivariate processes. Four empirical examples are included to show the effectiveness of the proposed method.

Keywords: Asymmetries; Matrix Inequality Constraints; Mixture Formulation; Multivariate Modelling; Optimal Forecasts; Power Transformations; Second Moment Structure

JEL Classification: C32, C53, C58, G15.

[*]Corresponding Author: Menelaos Karanasos, Economics and Finance, Brunel University London, UB3 3PH, UK; email: menelaos.karanasos@brunel.ac.uk, tel: +44(0)1895265284, fax: +44 (0)1895269770.

[⋆]Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK; email: xuy16@cardiff.ac.uk

1 Introduction

Multivariate GARCH/HEAVY[1] models, and MEM (multiplicative error models) have found considerable empirical success (see, for example, Nakatani and Teräsvirta, 2009, Shephard and Sheppard, 2010, Conrad and Karanasos, 2010, Noureldin et al., 2012, and Cipollini et al., 2013).

As was indicated at the outset of the research by Conrad and Karanasos (2010) the standard restricted extended multivariate GARCH model although allowing for spillovers requires all the parameters to be non-negative. As noted by Rabemananjara and Zakoian (1993) and Knight and Satchell (2007), parameters that take only non-negative values may be a source of important difficulties in running estimation procedures. If a shock in the past, regardless of the sign, always has a positive effect on the current conditional volatility, then the impact increases with the magnitude of the shock. Therefore, cyclical or any non-linear behavior in volatility cannot be taken into account. Equally important, and as pointed out by Cipollini and Gallo (2010), in our semi-unrestricted models, since they allow for negative conditional spillovers, the speed of absorption of a shock can be higher than in the restricted specification. In addition, the more parameters there are in the system (that is the higher its dimension) the more likely it is that one or more of them will take negative values (see, for example, in the empirical Section the estimation of high-low range volatilities in four equity markets). Thus, we need a mechanism to ensure that the conditional covariance matrix is positive definite almost surely at all points in time. Conrad and Karanasos (2010) obtained such a mechanism, but they presented explicit formulas only for the bivariate case of order $(1, 1)$.

In this paper, first, for the N-dimensional system of order $(1, q)$ we present a useful method for constructing tractable counterparts of the non-negativity constraints derived in Conrad and Karanasos (2010). The non-negativity (necessary and sufficient) conditions are easily modified, that is they are expressed in terms of matrix inequalities which can be solved easily. The research by Nelson and Cao (1992), He and Teräsvirta (1999), Gourieroux (2007), Tsai and Chan (2007, 2008), Nakatani and Teräsvirta (2008, 2009), Conrad (2010) and Conrad and Karanasos (2010), underlines the theoretical interest in the derivation of such necessary and sufficient conditions (this strand of the literature originated with the seminal work of Nelson, 1991). For example, one of our findings, that the jumps in realized volatility- as proxied by the high-low range volatility- have a negative impact on the realized volatility, is consistent with that in Andersen et al. (2007).

Our methodology is applicable to all three types of N-dimensional systems, that is GARCH/HEAVY models, and MEM. It is of considerable interest to investigate whether or not a number of reported estimated multivariate models satisfy these matrix inequality constraints. Indeed we find that an alarming number of seminal papers report estimated coefficients whose values violate the non-negativity conditions (see Section I of the supplementary Appendix).

Second, we also derive new tractable constraints for the asymmetric power versions of these N-dimensional systems. These allow new matrix inequalities to be constructed for the asymmetric power multivariate process and thus we extend the results in Conrad and Karanasos (2010). For example, our estimation of a trivariate MEM model produced four (out of nine) significant asymmetric parameters. In practice, these constraints are difficult to be satisfied. In other words, and as already noted earlier, they are commonly violated. Researchers should recognize that their existence might impose severe limitations on the parameter space. Thus, a critical question is: "If these non-negativity conditions are not fulfilled, is there an alternative multivariate model that allows for negative parameter values, which satisfy such constraints?". The answer is yes. One possibility is to employ a (unrestricted, since no constraints are imposed) multivariate exponential specification (see for example, Hautsch, 2008, and Taylor and Xu, 2017). However, it might be rather restrictive to use logs in all N cases.

Therefore, thirdly, and most importantly, we propose a new mixture formulation, which is an effective way to relax some of these constraints. In particular, we use the exponential function in some but not all of the N equations. By replacing power transformations with logarithmic ones we are cutting down the dimensions of the non-negativity conditions. These matrix inequality constraints are tractable in theory and practice. This general mixture formulation system includes the multivariate log-GARCH model (see Francq and Sucarrat, 2017, Francq et al., 2017) and exponential MEM (see, Hautsch, 2008) as special cases. For other recent developments in multivariate GARCH models see, for example, Pedersen and

[1] The acronym HEAVY (High frEquency bAsed VolatilitY) was introduced by Shephard and Sheppard (2010).

Rahbek (2014), Nielsen and Rahbek (2014), Morana (2015), and de Almeida et al. (2018).

Apart from conditions on multivariate GARCH models and MEM, which ensure the non-negativity of the conditional variables, we also derive (our fourth contribution) new theoretical results on the optimal forecasts and on the second order moments of such models, and, therefore, we extent the results in He and Teräsvirta (2004). He and Teräsvirta (1999), in the context of a univariate GARCH model and Conrad and Karanasos (2010) for a bivariate GARCH$(1,1)$ process, show that the less severe non-negativity conditions allow more flexibility in the shape of autocorrelation function than the constraints restricting the parameters to be non-negative. Similarly, allowing for negative values in some of the parameters should also improve the forecasting ability of the multivariate models. A Monte Carlo simulation forecasting exercise confirms this conjecture.

The relevance and the importance of the proposed method is demonstrated with four empirical examples on four different real datasets. Our matrix inequalities can be practically checked with ease and some of these even effortlessly enforced in estimation. A final contribution is the consistent estimation of the multivariate MEM by quasi maximum likelihood estimation instead of the efficient generalized method of moments estimation used in Cipollini et al. (2013). As an example, our conclusion that the conditional mean of stock volume has a negative impact on that of volatility (for two out of the five datasets used in the first empirical example; see Section 7), is in line with the theory by Wang (2007). According to Wang foreign purchases tend to lower volatility by increasing the investor base in emerging markets, since the broadening of the investor base improves the accuracy of market information and stabilizes stock prices.

The outline of the paper is as follows. Section 2 summarizes some basics concerning the notation used throughout the paper and introduces the vector asymmetric power specification. Section 3 reviews the symmetric model, and tractable expressions for the non-negativity constraints are presented together with some numerical examples. The next Section presents the Monte Carlo simulation. The main results in terms of matrix inequalities are stated in Sections 5 and 6. Section 7 contains the empirical examples, and the conclusions can be found in Section 8. The Appendix briefly discusses the optimal forecasts and the second moment structure of our proposed mixture formulation. A supplementary Appendix (available online) contains the proofs. Table 4 presents a summary of the various models. For a summary of our notation, see Table E.1 in Section E of the supplementary Appendix.

2 The Model

2.1 Notation

Throughout the paper we will adhere to the following notation. $\mathcal{F}_{t-1}^{(XF)}$ is the filtration generated by all available information through time $t-1$. We will use $\mathcal{F}_{t-1}^{(HF)}$ $(X=H)$ for the high frequency past data, i.e., for the case of the realized measure, or $\mathcal{F}_{t-1}^{(LoF)}$ $(X=Lo)$ for the low frequency past data, i.e., for the case of the close-to-close returns. Hereafter, for notational convenience we will drop the superscript XF.

We will use upper(lower) case boldface symbols to refer to square matrices(vectors). That is, $\mathbf{y}=[y_i]_{i=1,\dots,N}$ is an $N\times 1$ column vector, $\mathbf{Y}=[y_{ij}]_{i,j=1,\dots,N}$ is a square matrix of order N (hereafter we will drop the subscript for notational simplicity), and $diag[\mathbf{y}]$ denotes a diagonal matrix with elements y_i. Let also $Y_{ij}(L)$ be a polynomial of order N where L is the lag operator. Then $\mathbf{Y}(L)=[Y_{ij}(L)]$ indicates a matrix polynomial in the lag operator.

The identity matrix of order N is denoted by $\mathbf{I}_N$ and $\mathbf{0}_N$ is the null matrix (hereafter, we will drop the subscript if the order is N). $\mathbf{J}_{[d]}$, $0\leq d\leq N$, is a binary matrix that has ones in its first d rows and zeros elsewhere. Thus $\mathbf{J}_{[N]}$ or for notational convenience just $\mathbf{J}$ is the the unit matrix, that is a matrix with all its N^2 elements equal to 1. Alternatively, we can write it as $\mathbf{J}=\mathbf{j}\mathbf{j}'$ where $\mathbf{j}$ is a vector of ones. Similarly, $\mathbf{I}_{[d]}=\mathbf{I}\mathbf{J}_{[d]}$ is a diagonal matrix with ones in the first d diagonal elements and zeros elsewhere. Thus $\mathbf{I}_{(d)}=\mathbf{I}-\mathbf{I}_{[d]}$ is a diagonal matrix with ones in the last $N-d$ diagonal elements and zeros elsewhere. Clearly, $\mathbf{J}_{[0]}=\mathbf{I}_{[0]}=\mathbf{I}_{(N)}=\mathbf{0}$, and $\mathbf{I}_{[N]}=\mathbf{I}_{(0)}=\mathbf{I}$.

Further, using standard notation, $\mathbf{Y}'$ and $\mathbf{Y}^{-1}$ are the transpose and the inverse of the square matrix $\mathbf{Y}$. The determinant and the adjoint of $\mathbf{Y}$ are denoted by $\det[\mathbf{Y}]$ and $adj[\mathbf{Y}]$, respectively. That is, $adj[\mathbf{Y}]=[Y_{ij}^{\{a\}}]$ with $Y_{ij}^{\{a\}}=(-1)^{i+j}\det[\mathbf{Y}_{\{ji\}}]$ where $\mathbf{Y}_{\{ji\}}$ is the $\mathbf{Y}$ matrix without its jth row and ith column. In other words, $Y_{ij}^{\{a\}}$ is the cofactor of the jith element of $\mathbf{Y}$.

In addition, the elementwise expectation operator is denoted by $\mathbb{E}$, i.e., $\mathbb{E}(\mathbf{Y}) = [\mathbb{E}(y_{ij})]$ (similarly, $\mathbb{E}(\mathbf{Y}\,|\mathcal{F}_{t-1})$ denotes the elementwise, conditional on time $t-1$, expectation operator), whereas $\mathbf{Y}^{\wedge k} = [y_{ij}^k]$ is the elementwise exponentiation. We will refer to the elementwise absolute value of $\mathbf{Y}$ as $abs[\mathbf{Y}] = [|y_{ij}|]$. The inequality $\mathbf{Y} \geq \mathbf{0}$ means that all elements of $\mathbf{Y}$ are non-negative real numbers.

Moreover, $\max[\mathbf{Y}]$ indicates the largest element of the matrix $\mathbf{Y}$ and $\mathbf{Y}^k = \prod_{i=1}^k \mathbf{Y}$ means that the matrix $\mathbf{Y}$ is raised to the power of k. Finally, let $\mathbf{Y}^{\otimes 2} = \mathbf{Y} \otimes \mathbf{Y}$, $\mathbf{Y}^{\otimes I} = \mathbf{Y} \otimes \mathbf{I}$, $\mathbf{I}^{\otimes Y} = \mathbf{I} \otimes \mathbf{Y}$ where $\otimes$ is the Kronecker product of two matrices, $vec(\mathbf{Y})$ is a vector in which the columns of the matrix $\mathbf{Y}$ are stacked one underneath the other, and $\odot$ is the Hadamard or elementwise product of matrices.

2.2 The Asymmetric Power Specification

In this Section we introduce the asymmetric power specification. Consider the N-dimensional vector process, $\boldsymbol{\varepsilon}_t = [|\varepsilon_{it}|^{\delta_i}]$ where $\delta_i \in \mathbb{R}^+$. We assume that the vector $\boldsymbol{\varepsilon}_t$ is characterized by the relation

$$\boldsymbol{\varepsilon}_t = \mathbf{Z}_t \boldsymbol{\sigma}_t, \tag{1}$$

where $\mathbf{Z}_t = diag[\mathbf{z}_t]$, $\mathbf{z}_t = [|e_{it}|^{\delta_i}]$, and $\boldsymbol{\sigma}_t = [\sigma_{it}^{\delta_i}]$ is $\mathcal{F}_{t-1}$ measurable with $\mathcal{F}_{t-1} = \sigma(\boldsymbol{\varepsilon}_{t-1}, \boldsymbol{\varepsilon}_{t-2}, \ldots)$. That is, $\boldsymbol{\varepsilon}_t = [|e_{it}|^{\delta_i}\, \sigma_{it}^{\delta_i}]$.

Let also $\widetilde{\boldsymbol{\sigma}}_t = [\sigma_{it}]$, in other words, $\widetilde{\boldsymbol{\sigma}}_t$ is equal to $\boldsymbol{\sigma}_t$ when $\delta_i = 1$ for all i. Further, let $\mathbf{E}_t = diag[\mathbf{e}_t]$ where the stochastic vector $\mathbf{e}_t = [e_{it}]$ is independent and identically distributed ($i.i.d$). In addition, let $\widetilde{\boldsymbol{\varepsilon}}_t = \mathbf{E}_t \widetilde{\boldsymbol{\sigma}}_t = [e_{it}\sigma_{it}]$.

In the N-dimensional GARCH model $\mathbf{e}_t$ has zero mean zero, unit variance, and positive definite time invariant correlation matrix $\mathbf{R} = [\rho_{ij}]$ with $\rho_{ii} = 1$- notice that in this case the ith element of $\mathbf{e}_t$ is equal to the corresponding element of $\mathbf{z}_t$, when $\delta_i = 1$ for all i, multiplied by $sign(e_{it})$- therefore, $\widetilde{\boldsymbol{\varepsilon}}_t$ is a vector with zero conditional mean: $\mathbb{E}(\widetilde{\boldsymbol{\varepsilon}}_t\,|\mathcal{F}_{t-1}) = \mathbf{0}$. The conditional covariance matrix of $\widetilde{\boldsymbol{\varepsilon}}_t$ is given by $\boldsymbol{\Sigma}_t = \mathbb{E}(\widetilde{\boldsymbol{\varepsilon}}_t \widetilde{\boldsymbol{\varepsilon}}_t{}'\,|\mathcal{F}_{t-1}) = diag[\widetilde{\boldsymbol{\sigma}}_t]\mathbf{R}diag[\widetilde{\boldsymbol{\sigma}}_t]$.

In the N-dimensional MEM $\mathbf{e}_t > \mathbf{0}$, with $\mathbb{E}(\mathbf{e}_t) = \mathbf{j}$, and positive definite covariance matrix $\mathbf{Q} = [q_{ij}]$, with $\mathbf{q} = diag[\mathbf{Q}]$. That is, $\mathbb{E}(\widetilde{\boldsymbol{\varepsilon}}_t\,|\mathcal{F}_{t-1}) = \widetilde{\boldsymbol{\sigma}}_t$. In this case $\boldsymbol{\Sigma}_t = \mathbb{E}(\widetilde{\boldsymbol{\varepsilon}}_t \widetilde{\boldsymbol{\varepsilon}}_t{}'\,|\mathcal{F}_{t-1}) = diag[\widetilde{\boldsymbol{\sigma}}_t]\mathbf{Q}diag[\widetilde{\boldsymbol{\sigma}}_t]$.

As pointed out by Conrad and Karanasos (2010) a major problem in specifying a valid multivariate GARCH process or MEM lies in choosing appropriate parametric specifications for $\widetilde{\boldsymbol{\sigma}}_t^{\wedge 2}$ such that $\boldsymbol{\Sigma}_t$ is positive definite almost surely for all t. Positive definiteness of $\boldsymbol{\Sigma}_t$ (for all t) follows if, in addition to the correlation matrix $\mathbf{R}$ being positive definite, the conditional variances of ε_{it} or their power transformations, $\sigma_{it}^{\delta_i}$, are positive (for all t) as well. Similarly, for the MEM we need in addition to the covariance matrix $\mathbf{Q}$ being positive definite, the conditional means of ε_{it} or their power transformations, $\sigma_{it}^{\delta_i}$, to be positive for all t.

The N-dimensional semi-unrestricted extended asymmetric power (SUE-AP) model of order $(1,1)$ -in what follows for notational simplicity we will drop the order of the model if it is $(1,1)$- consists of the following equations:

$$\sigma_{it}^{\delta_i} = \omega_i + \sum_{j=1}^N (\alpha_{ij} + \gamma_{ij} s_{j,t-1})\,|\varepsilon_{j,t-1}|^{\delta_j} + \sum_{j=1}^N \beta_{ij}\sigma_{j,t-1}^{\delta_j},$$

where s_{jt} is a dummy variable that; i) in the case of the GARCH model takes the value 1 if $e_{jt} < 0$, 0 otherwise, that is $s_{jt} = 0.5[1 - sign(e_{jt})1]$, and ii) takes the value 1 if a signed variable (i.e., stock returns) $x_{jt} < 0$, 0 otherwise.[2]

This can be either a multivariate GARCH/HEAVY model or a MEM. For example, in the bivariate context the two GARCH variables, ε_{it}, $i = 1, 2$, can be the stock returns and the signed square rooted (SSR) realized measure (i.e., realized variance), and $\sigma_{it} = E(\varepsilon_{it}^2\,|\mathcal{F}_{t-1})$ their conditional variances. The HEAVY formulation parallels the GARCH one. It is also very similar to the bivariate MEM. In the latter model the two variables (ε_{it}) can be the squared returns and the realized measure, and $\sigma_{it} = E(\varepsilon_{it}\,|\mathcal{F}_{t-1})$ their conditional means. Therefore we will use the three terms, GARCH, HEAVY, MEM, interchangeably.

[2] This type of asymmetry was introduced by Glosten et. al. (1993). In Section B of the supplementary Appendix we will consider a second type of asymmetry, which was introduced by Ding et. al. (1993), and we will refer to it as SUE-AP model 2.

The SUE-AP model can be expressed/interpreted as an N-dimensional system with shock (unconditional) and conditional spillovers:

$$(\mathbf{I}-\mathbf{B}L)\boldsymbol{\sigma}_t = \boldsymbol{\omega} + L\mathbf{A}_t\boldsymbol{\varepsilon}_t, \tag{2}$$

(recall that $\boldsymbol{\varepsilon}_t$ and $\boldsymbol{\sigma}_t$ have been defined in eq. (1))where $\mathbf{B} = [\beta_{ij}]$ is a full matrix (of order N) that its cross diagonal elements capture the conditional spillovers; $\boldsymbol{\omega} = [\omega_i]$ is a vector that contains the drifts; $\mathbf{A}_t = \mathbf{A} + \mathbf{\Gamma}_t$, where $\mathbf{A} = [\alpha_{ij}]$ and $\mathbf{\Gamma}_t = [\boldsymbol{\gamma}_{ij} s_{jt}]$, are full matrices as well. Note that $\mathbf{\Gamma}_t$ can be written as $\mathbf{\Gamma}_t = \mathbf{\Gamma} diag[\mathbf{s}_t]$ where $\mathbf{\Gamma} = [\gamma_{ij}]$ and $\mathbf{s}_t = [s_{it}]$. The cross diagonal elements of $(\mathbf{\Gamma}_t)\mathbf{A}$ capture the (asymmetric) shock (or unconditional) spillovers.

The above model is termed extended since the square matrices are full, and semi-unrestricted because as we will see below some of the elements of the $\mathbf{B}$ matrix (including some of the off-diagonal ones but not all) are allowed to take not only positive but negative values as well. That is, as in Conrad and Karanasos (2010) we consider a SUE formulation that allows for feedback effects between the conditional variances (or means in the case of the MEM), which can be of either sign, positive or negative.[3] However, on the other hand all the elements of the $\mathbf{A}$ and $\mathbf{\Gamma}$ matrices are restricted to be non-negative (see Theorem 2 below). Therefore, in Section 5 we will also introduce another model which allows for negative (asymmetric as well) unconditional spillovers either own or cross ones. Note that, if there are no asymmetries, that is $\mathbf{\Gamma} = 0$, then the model reduces to the symmetric power one:

$$(\mathbf{I}-\mathbf{B}L)\boldsymbol{\sigma}_t = \boldsymbol{\omega} + \mathbf{A}L\boldsymbol{\varepsilon}_t, \tag{3}$$

which is further reduced to the benchmark GARCH model examined in Conrad and Karanasos (2010) if $\delta_i = 2$ for all i:

$$(\mathbf{I}-\mathbf{B}L)\widetilde{\boldsymbol{\sigma}}_t^{\wedge 2} = \boldsymbol{\omega} + \mathbf{A}L\widetilde{\boldsymbol{\varepsilon}}_t^{\wedge 2},$$

or the benchmark MEM if $\delta_i = 1$ for all i:

$$(\mathbf{I}-\mathbf{B}L)\widetilde{\boldsymbol{\sigma}}_t = \boldsymbol{\omega} + \mathbf{A}L\widetilde{\boldsymbol{\varepsilon}}_t.$$

As already noted above a crucial problem concerns the identification of necessary and sufficient conditions for the SUE model to have positive $\boldsymbol{\sigma}_t$ for all t. This will be the topic of analysis in the rest of the paper.

3 Matrix Inequality Constraints

To keep this article relatively self-contained we briefly review the main theoretical results of Conrad and Karanasos (2010) on the derivation of necessary and sufficient conditions, which ensure that $\boldsymbol{\sigma}_t$ in an N-dimensional symmetric GARCH system (or MEM as well) is positive almost surely for all t. In this Section we give an outline of the second main step in the derivation of such (necessary and sufficient) conditions. The first step, that is the 'univariate' representations, which each conditional variance (or mean in the case of the MEM) admits, is given in Section A of the supplementary Appendix as Lemma A1. The second step is the infinite expansions, in terms of convolutions of infinite-order kernels and corresponding power transformed errors, of the aforementioned 'univariate' representations. The latter two steps constitute the main steps in the derivation of Theorem 1 in Conrad and Karanasos (2010), which we state as Proposition A1 in Section A of the supplementary Appendix.

After the two main steps in the derivation of the aforementioned proposition, which are presented for completeness, then our more general result follows on formal grounds. That is we express the non-negativity constraints for the symmetric system as matrix inequalities (Theorem 1 below). A natural extension of this theorem is the generalization of the results to the asymmetric case. This is developed in Section 5.

[3] In the symmetric restricted extended formulation (see Jeantheau, 1998 and Ling and McAleer, 2003) the $\mathbf{A}$ and $\mathbf{B}$ are full matrices but all their elements are allowed to take only non-negative values. As pointed out by Conrad and Karanasos (2010) the assumption that only positive feedback is allowed for is tempting because positive constants and parameter matrices with non-negative coefficients are a sufficient condition for the positive definiteness of the conditional covariance matrix in the extendend formulation.

3.1 Wold Decompositions

In this Section we will introduce a useful lemma. In particular, we obtain the SUE-P infinite-order expansion of each power transformed conditional variable in terms of convolutions of GARCH/MEM kernels and corresponding power transformed errors.

First, some additional notation is needed. Set $\boldsymbol{\mu} = [\mu_i]$ as

$$\boldsymbol{\mu} = adj[\mathbf{I} - \mathbf{B}]\boldsymbol{\omega}. \tag{4}$$

To ease the following explanations let $\beta(L) = 1 - \sum_{i=1}^{N} \beta_i L^i = \prod_{i=1}^{N} (1 - \phi_i L)$ be

$$\beta(L) = \det[\mathbf{I} - \mathbf{B}L], \tag{5}$$

which, since under the assumptions in Appendix A: $\beta_N \neq 0$, is a scalar polynomial of order N; ϕ_i are the roots of $\beta(z^{-1})$. In what follows, without loss of generality, we will assume that they are distinct and they satisfy the inequalities: $|\phi_1| > |\phi_2| > \cdots > |\phi_N|$; see also Assumption A2 in Appendix A. Next we define the square matrix polynomial $\boldsymbol{\alpha}(L) = [a_{ij}(L)]$ with $a_{ij}(L) = \sum_{n=1}^{N} a_{ij}^{(n)} L^n$ (where the superscript with parenthesis denotes an index):

$$\boldsymbol{\alpha}(L) = adj[\mathbf{I} - \mathbf{B}L]\mathbf{A}L. \tag{6}$$

Similarly to $\beta(L)$, since under the assumptions in Appendix A: $a_{ij}^{(N)} \neq 0$ for all i, j, the scalar polynomials $a_{ij}(L)$ are of order N.

Next, set $\boldsymbol{\Psi}(L) = [\Psi_{ij}(L)]$, where $\Psi_{ij}(L) = \sum_{k=1}^{\infty} \psi_{ij}^{(k)} L^k$, as

$$\boldsymbol{\Psi}(L) = \boldsymbol{\alpha}(L)/\beta(L), \tag{7}$$

and thus

$$\Psi_{ij}(L) = \alpha_{ij}(L)/\beta(L). \tag{7a}$$

Equivalently $\boldsymbol{\Psi}(L)$ can be written as

$$\boldsymbol{\Psi}(L) = \sum_{k=1}^{\infty} \boldsymbol{\Psi}_k L^k, \tag{7b}$$

with $\boldsymbol{\Psi}_k = [\psi_{ij}^{(k)}]$.

The following corollary gives the one-sided representation of the SUE-P model. The proof is trivial (the result is obtained from eq. (3) by inverting the matrix polynomial $\mathbf{I} - \mathbf{B}L$ using eq. (5)).

Corollary 1 *Let Assumptions (A1) and (A2) in Appendix A be satisfied. Then the SUE-P model in eq. (3) admits the multivariate Wold decomposition:*

$$\boldsymbol{\sigma}_t = \frac{\boldsymbol{\mu}}{\beta(1)} + \frac{\boldsymbol{\alpha}(L)}{\beta(L)}\boldsymbol{\varepsilon}_t, \tag{8}$$

with the corresponding 'univariate' one-sided representations given by

$$\sigma_{it}^{\delta_i} = \frac{\mu_i}{\beta(1)} + \sum_{j=1}^{N} \Psi_{ij}(L) \left|\varepsilon_{jt}\right|^{\delta_j}. \tag{8a}$$

(see also Lemma 2 in Conrad and Karanasos, 2010).

Here, each $\Psi_{ij}(L)$ (see eq. (7a)) can be thought of as an infinite-order kernel of a univariate SUE-P(N, N) model. Clearly, for the N-dimensional process in eq. (3) to be well-defined and the N (powered transformed) conditional variables to be positive almost surely for all t, all the constants μ_i (see eq. (4)) must be positive and all the $\psi_{ij}^{(k)}$ coefficients in the 'univariate' Wold decompositions, that is eq. (8a), must be non-negative: $\psi_{ij}^{(k)} \geq 0$, $i, j = 1, \ldots, N$ for $k = 1, 2, \ldots$

In other words, the non-negativity of the (powered transformed) conditional variables is guaranteed if and only if all the kernels are non-negative, i.e., if the *infinite* number of coefficients in the one-sided expansions of the N^2 kernels are non-negative. For this, one should express these coefficients as functions of the parameters of the original process. It can then be shown that checking a *finite* number of inequality constraints on these parameters ensures the non-negativity of all GARCH/MEM kernels of the SUE-P model (see Conrad and Karanasos, 2010, who paid special attention only to the bivariate case of order $(1,1)$).

Alternative One-Sided Representation

In this Section we make a general observation that will be applied tactically later on. That is, we shall make use of the following corollary. The multivariate Wold decomposition has been presented above. Here we present an alternative form for such an infinite-order expansion.

Corollary 2 *Let Assumptions (A1) and (A2) in the Appendix A be satisfied. Then eq. (8) can be rewritten in an alternative form as:*

$$\boldsymbol{\sigma}_t = \frac{\boldsymbol{\mu}}{\beta(1)} + \sum_{k=1}^{\infty} \mathbf{B}^{k-1}\mathbf{A}L^k\boldsymbol{\varepsilon}_t.$$

The above corollary follows directly from Corollary 1 since $(\mathbf{I}-\mathbf{B}L)^{-1}\boldsymbol{\omega} = \frac{adj[\mathbf{I}-\mathbf{B}]\boldsymbol{\omega}}{\det[\mathbf{I}-\mathbf{B}]} = \frac{\boldsymbol{\mu}}{\beta(1)}$ and $\boldsymbol{\Psi}(L) = \frac{\boldsymbol{\alpha}(L)}{\beta(L)} = \frac{adj[\mathbf{I}-\mathbf{B}L]\mathbf{A}L}{\beta(L)} = [\mathbf{I}-\mathbf{B}L]^{-1}\mathbf{A}L = \sum_{k=1}^{\infty}\mathbf{B}^{k-1}\mathbf{A}L^k$ and hence $\boldsymbol{\Psi}_k = \mathbf{B}^{k-1}\mathbf{A}$.

3.2 Tractable Expressions

In this Section we will show that the non-negativity conditions in Conrad and Karanasos (2010) can be expressed as simple inequalities involving square matrices. Our constraints in terms of these inequalities are algorithmically solvable fast enough to be practically relevant. In other words, the result of this Section makes the problem of non-negativity conditions for N-dimensional SUE-P systems easily solvable and downright tractable.

Next we will present in the following theorem our tractable non-negativity constraints. As we have just noted they are expressed in terms of matrix inequalities, which can be easily computed fast enough to make them practical. But before we do that we will introduce some further notation.

If $\mathbf{Y}^{(n)} = [y_{ij}^{(n)}]$, $n = 1,\ldots,N$ (recall that the superscript with parenthesis denotes an index), then $\max[\mathbf{Y}^{(n)}] = \max(y_{ij}^{(n)})$ for $i,j = 1,\ldots,N$. In other words, $\max[\mathbf{Y}^{(n)}]$ is the largest element of all the N^2 elements of $\mathbf{Y}^{(n)}$. In addition, $\mathbf{Y}_{\max}^{(n)} = [\max_{1\leq n\leq N} y_{ij}^{(n)}]$ is a matrix whose ijth element is the largest of the N $y_{ij}^{(n)}$ elements. The following theorem holds (below $\log[\mathbf{X}]$ means that we take the log of each element of $\mathbf{X}$).

Theorem 1 *Consider the N-dimensional vector SUE-P model in eq. (3) and let Assumptions (A1)-(A2) in Appendix A be satisfied. Then, the necessary and sufficient conditions for $\sigma_{it}^{\delta_i} > 0$, $i = 1,\ldots,N$, for all t can be expressed as:*

(a) $\boldsymbol{\mu} = adj[\mathbf{I}-\mathbf{B}]\boldsymbol{\omega} > \mathbf{0}$

$$(b)\begin{cases} \phi_1 \text{ is real, and } \phi_1 > 0, & (C1)\\ adj[\mathbf{I}\phi_1 - \mathbf{B}]\mathbf{A} > \mathbf{0}, & (C2)\\ \mathbf{B}^{k-1}\mathbf{A} \overset{(k\leq\kappa_{ij})}{\succeq} \mathbf{0}, \text{ for } k = 1,\ldots,\kappa, & (C3)\end{cases}$$

(the symbol $\overset{(k\leq\kappa_{ij})}{\succeq}$ means that if $k > \kappa_{ij}$ then the ijth scalar inequality of the matrix inequality $\mathbf{B}^{k-1}\mathbf{A} \geq \mathbf{0}$ should be disregarded)[4]

[4]Let the ijth element of $\mathbf{B}^{k_{ij}}$ be denoted by $\beta_{ij}^{[k_{ij}]}$, $i,j = 1,\ldots,N$, and construct the matrix $\mathbf{B}^{[k_{ij}]} = [\beta_{ij}^{[k_{ij}]}]$. Then condition (C3***) in Proposition A1 in the supplementary Appendix for $k_{ij} = 1,\ldots,\kappa_{ij}$ amounts to: $\mathbf{B}^{[k_{ij}]}\mathbf{A} > \mathbf{0}$ for all $i,j = 1,\ldots,N$. For notational convenience we write it instead as in Condition (C3): $\mathbf{B}^{k-1}\mathbf{A} \overset{(k\leq\kappa_{ij})}{\succeq} \mathbf{0}$.

(b^*) The κ_{ij} and κ in Condition (C3) are obtained as follows. Let $\kappa = \max[\mathbf{K}]$, $\mathbf{K} = [\kappa_{ij}]$, and κ_{ij} is the smallest integer greater than or equal to $\max\{0, \varphi_{ij}\}$ *with* $\mathbf{\Phi} = [\varphi_{ij}]$*, given by*

$$\mathbf{\Phi} = [\log[\mathbf{H}^{(1)}] - \log[(N-1)]\mathbf{H}^{(n)}_{\max}][\log(|\phi_2|) - \log(|\phi_1|)]^{-1},$$

where $\mathbf{H}^{(n)}_{\max}=[\max_{2\leq n\leq N}\eta^{(n)}_{ij}]$ *and* $\mathbf{H}^{(n)} = [\eta^{(n)}_{ij}]$, $1 \leq n \leq N$*, is given by*

$$\mathbf{H}^{(n)} = abs\left[\frac{adj[\mathbf{I}\phi_n - \mathbf{B}]\mathbf{A}}{\sum_{j=1}^{N} j\beta_j\phi_n^{N-(j-1)}}\right].$$

Theorem 1 follows directly from Proposition A1 in Section A of the supplementary Appendix (see also Theorem 1 in Conrad and Karanasos, 2010) and Corollary 2. Interestingly, we only have to check: i) from condition (a) if all the N elements of the vector $adj[\mathbf{I}-\mathbf{B}]\boldsymbol{\omega}$ are positive, ii) from Condition (C2) if all the N^2 elements of the matrix $adj[\mathbf{I}\phi_1 - \mathbf{B}]\mathbf{A}$ are positive, and iii) from Condition (C3) if all the N^2 elements of each of the k matrices, $\mathbf{B}^{k-1}\mathbf{A}$, are non-negative (clearly, the latter condition, for $k = 1$, implies that $\mathbf{A} \geq \mathbf{0}$).

In other words, we replace $\alpha_{ij}(\phi_1^{-1}) > 0$, for $i, j = 1, \ldots, N$, in Proposition A1 by its equivalent matrix expression $adj[\mathbf{I}\phi_1 - \mathbf{B}]\mathbf{A} > \mathbf{0}$, and likewise $\psi^{(k)}_{ij} \geq 0$, for $i, j = 1, \ldots, N$ by its equivalent matrix expressions, $\mathbf{\Psi}_k = \mathbf{B}^{k-1}\mathbf{A} \geq \mathbf{0}$ for all k (see Condition (C3) in part (b) of the above Theorem and also its part (b^*).

However, in practice one should just check the sufficient conditions: $\mathbf{B}^{k-1}\mathbf{A} \geq \mathbf{0}$ for all k from 1 up to a large enough k, i.e., $k = N$. In other words, in practice part (b^*) of the above theorem can be disregarded and Condition (C3) simplifies considerably since it reduces to: $\mathbf{B}^{k-1}\mathbf{A} \geq \mathbf{0}$, for $k = 1, \ldots, \kappa$ with κ, for example, equal to N. This simplification of the above Theorem is another important consequence of our matrix inequality constraints. Therefore, our conditions are algorithmically solvable fast enough to be practically relevant. It is very easy for the practitioner to check if these matrix inequality constraints are satisfied.

Next we show that the matrix inequalities are easily represented in terms of scalar inequalities as well. As a last stage before we do that, however, we will introduce some additional notation. First, let $\mathbf{B}^\star = [\beta^\star_{ij}]$ be given by $\mathbf{B}^\star = \mathbf{I} - \mathbf{B}$, then condition (a) in Theorem 1 implies that:

$$\sum_{m=1}^{N} \beta^{\star\{a\}}_{im}\omega_m > 0, \text{ for all } i = 1, \ldots, N, \text{ (a}')$$

where $\beta^{\star\{a\}}_{im} = (-1)^{i+m}\det[\mathbf{B}^\star_{\{mi\}}]$, and the latter matrix is obtained by deleting the mth row and the ith column from $\mathbf{B}^\star$. Similarly, let $\mathbf{B}^* = [\beta^*_{ij}]$ be given by $\mathbf{I}\phi_1 - \mathbf{B}$, then Condition (C2) in Theorem 1 is equivalent to:

$$\sum_{m=1}^{N} \beta^{*\{a\}}_{im}\alpha_{mj} > 0, \text{ for all } i, j = 1, \ldots, N \text{ (C2}'),$$

where $\beta^{*\{a\}}_{im} = (-1)^{i+m}\det[\mathbf{B}^*_{\{mi\}}]$. Next, let $\boldsymbol{\phi} = [\phi_i]$ be the vector of the N distinct roots (see eq. (5)). Then there is a nonsingular matrix $\mathbf{\Lambda} = [\lambda_{ij}]$ (the matrix with the N eigenvectors of $\mathbf{B}$) such that

$$\mathbf{B}^k = \mathbf{\Lambda} diag[\boldsymbol{\phi}^{\wedge k}]\mathbf{\Lambda}^{-1}.$$

Denote the ijth element of $\mathbf{\Lambda}^{-1}$ by λ^*_{ij}, that is $\mathbf{\Lambda}^{-1} = [\lambda^*_{ij}]$. Then, for each k, Condition (C3) in Theorem 1, that the ijth element of $\mathbf{\Psi}^{(k)} = \mathbf{B}^{k-1}\mathbf{A}$ must be non-negative for all i, j, amounts to:

$$\psi^{(k)}_{ij} = \sum_{m=1}^{N}\sum_{l=1}^{N} \lambda_{il}\lambda^*_{lm}\phi_l^{k-1}\alpha_{mj} > 0, \text{ for all } i, j = 1, \ldots, N, \text{ (C3}').$$

The above inequalities when $k = 1$ reduce to: $\alpha_{ij} > 0$, since $\sum_{l=1}^{N} \lambda_{il}\lambda^*_{lm} = 1$ if $i = m$, and zero otherwise.

3.3 Trivariate System and Numerical Examples

In this Section numerical examples are included to show the effectiveness of the proposed method. These may be helpful to the researcher who wishes to skip theoretical derivations and is mainly interested in the application of these constraints to a given N-dimensional system at hand. Next we will discuss a specific model in order to make our analysis more concise. That is, for illustrative purposes, we will consider the trivariate case.

Lemma 1 *Let Assumptions (A1) and (A2) in Appendix A be satisfied. The following conditions are necessary and sufficient for* $\sigma_{it}^{\delta_i} > 0$, $i = 1, 2, 3$, *for all* t, *in the trivariate SUE-P model (with* $\kappa = 3$*):*
(a) For the three constants we require

$$\sum_{m=1}^{3} \beta_{im}^{\star\{a\}} \omega_m > 0, \text{ for all } i = 1, 2, 3 \text{ where}$$
$$\beta_{im}^{\star\{a\}} = \begin{cases} (1-\beta_{ll})(1-\beta_{nn}) - \beta_{ln}\beta_{nl} & \text{if} \quad i = m;\ l \neq n \neq i, \\ (-1)^{i+m+1}[\beta_{im}(1-\beta_{ll}) + \beta_{il}\beta_{lm}] & \text{if} \quad i \neq m \neq l \end{cases}$$

$$(b) \begin{cases} \phi_1 \text{ is real, and } \phi_1 > 0, & (C1) \\ \sum_{m=1}^{3} \beta_{im}^{*\{a\}} \alpha_{mj} > 0, \text{ for all } i, j = 1, 2, 3, \text{ where} & \\ \beta_{im}^{*\{a\}} = \begin{cases} (\phi_1-\beta_{ll})(\phi_1-\beta_{nn}) - \beta_{ln}\beta_{nl} & \text{if} \quad i = m,\ l \neq n \neq i, \\ (-1)^{i+m+1}[\beta_{im}(\phi_1-\beta_{ll}) - \beta_{il}\beta_{lm}] & \text{if} \quad i \neq m \neq l, \end{cases} & (C2') \\ a_{ij} \geq 0, \ \sum_{m=1}^{3} \beta_{im}\alpha_{mj} \geq 0, \text{ and } \sum_{m=1}^{3}\sum_{l=1}^{3} \beta_{il}\beta_{lm}\alpha_{mj} > 0, \text{ for all } i, j = 1, 2, 3. & (C3') \end{cases}$$

In what follows we graphically illustrate the necessary and sufficient parameter set for the trivariate SUE-P system. This will provide a better understanding of the results presented in the previous Subsection. We discuss four examples. We allow two off-diagonal elements of $\mathbf{B}$ to vary from -0.5 to 0.5. In the first example, we examine the situation where b_{13} and b_{31} vary. The purpose is to see if bidirectional negative (conditional) spillovers are permitted. In the second example, we allow b_{21} and b_{31} (i.e., two parameters in the first column of $\mathbf{B}$) to vary. The purpose is to see if negative (conditional) spillovers from one variable to the other two variables can be allowed. In the third example, we vary b_{21} and b_{23} (i.e., two parameters in the second row of $\mathbf{B}$). The purpose is to see if negative spillovers from two conditional variables to the third one can be allowed. In the fourth example, we examine if more than two off-diagonal elements of the $\mathbf{B}$ matrix can be negative. To do so, we restrict b_{21} to be negative and vary b_{13} and b_{31}.

The parameters chosen are mainly from the empirical results in Table 6 presented in Section 7 (in particular, the results from the FTSE index in dataset 2). The four data generating processes (DGP) are given by:

Table 1A. Data generating process for Examples 1 and 2.

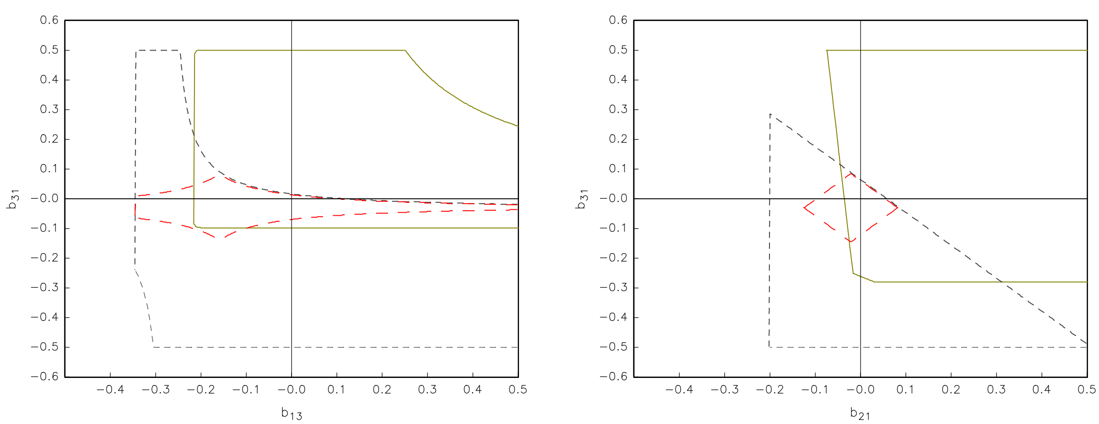
	DGP Ex.1			DGP Ex.2		
$\boldsymbol{\omega}'$	0.149	0.074	0.124	0.153	0.106	0.103
$\mathbf{A}$	0.064	0.021	0.158	0.075	0.017	0.123
	0.008	0.005	0.108	0.021	0.002	0.109
	0.028	0.043	0.198	0.030	0.037	0.104
$\mathbf{B}$	0.790	0.032	b_{13}	0.780	0.003	0.017
	0.001	0.808	0.006	b_{21}	0.901	0.022
	b_{31}	0.137	0.616	b_{31}	0.082	0.650

Table 1B. Data generating process for Examples 3 and 4.

	DGP Ex.3			DGP Ex.4		
$\boldsymbol{\omega}'$	0.162	0.114	0.117	0.214	0.184	0.164
A	0.075	0.011	0.140	0.078	0.012	0.171
	0.013	0.003	0.139	0.012	0.005	0.100
	0.023	0.044	0.201	0.048	0.029	0.228
B	0.744	0.002	0.051	0.743	0.031	b_{13}
	b_{21}	0.901	b_{23}	$-\mathbf{0.028}$	0.851	0.053
	0.009	0.056	0.559	b_{31}	0.111	0.548

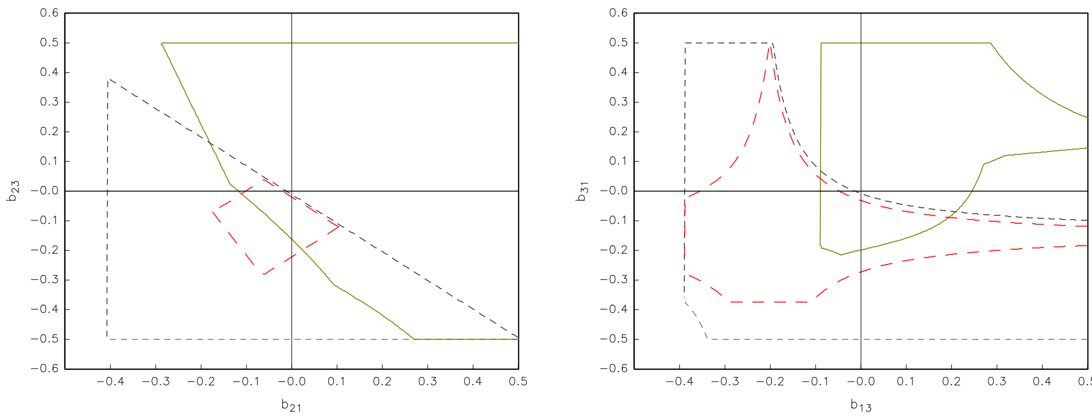
In the following figures, the lines show which combinations of the two free parameters satisfy the necessary and sufficient conditions of Theorem 1 and those for the existence of the first and second unconditional moments (reported in Appendices B and C, respectively). We begin by discussing the implications of Example 1, which is presented in Figure 1a. First, all combinations of b_{13} and b_{31} that are bounded by the bold solid lines satisfy the conditions of Theorem 1. Second, the combinations of the two parameters, which are bounded by the dotted grey (dashed red) lines, satisfy the conditions for the existence of the first (second) unconditional moments. Interestingly, both off-diagonal elements can be negative simultaneously.

Example 2 is visualized in Figure 1b. The conditions of Theorem 1 allow for negative spillovers from $\sigma_{1t}^{\delta_1}$ to $\sigma_{2t}^{\delta_2}$ and $\sigma_{3t}^{\delta_3}$. The negative parameter set that satisfies all conditions simultaneously is given by the area that is above and to the right of all the three lines in the third quadrant. Figure 1c shows that, for the parameters in Example 3, the conditions of Theorem 1 allow for negative spillovers from $\sigma_{1t}^{\delta_1}$ and $\sigma_{3t}^{\delta_3}$ to $\sigma_{2t}^{\delta_2}$. From example 4, it is interesting to observe that three off-diagonal elements in the **B** matrix can be negative and, at the same time satisfy all the non-negativity conditions.



a. Example 1

b. Example 2



c. Example 3 d. Example 4

FIGURE 1. Necessary and sufficient parameter sets for the trivariate SUE-P system from Examples 1 to 4. Solid brown lines represent the restrictions implied by Theorem 1. Dotted grey lines represent the restrictions implied by the existence of the unconditional first moment. Dashed red lines represent the restrictions implied by the existence of the unconditional second moment.

4 Monte Carlo Simulations

In this Section we employ the symmetric SUE model in eq. (3) with $\delta_i = 1$ for all i, and by using Monte Carlo simulations, we examine the effects of ignoring the non-negativity conditions in Theorem 1 on: i) the bias of quasi maximum likelihood (QML) estimates, and ii) the out of sample forecasts. We compare three cases: I) imposing our matrix inequality constraints in the estimation, II) enforcing Bollerlsev's sufficient conditions, and III) the unconstrained estimation.

The DGP in the context of MEM is as follows. To generate the disturbance vector $\mathbf{e}_t$, we use the multivariate log-normal distribution, with unit vector $\mathbf{j}$ as an expectation and covariance matrix $\mathbf{Q} = [q_{ij}]$ (using dataset 2, the FTSE index; see the empirical Section and Table 6 below). The results are based on 1000-repetition Monte Carlo simulations each with a sample size of 1000 observations. It should be noted that there should be no differences between the three alternative estimates if all the parameter values in the DGP are non-negative.

In our DGP, we set three elements in the $\mathbf{B}$ matrix to be negative (b_{13}, b_{21}, b_{31}), but still maintain the matrix inequality constraints in Theorem 1. The parameter values are reported in the first column of Table 2. The estimates based on the matrix inequality constraints have smaller bias than the other two. The performance of the estimates without imposing any non-negativity conditions is the worst both in terms of the bias and the standard deviation.

Table 2. The mean and standard deviation of QML estimates.

		Mean			Std.		
	True	Case I	Case II	Case III	Case I	Case II	Case III
ω_{11}	0.214	0.244	0.185	0.245	0.220	0.111	0.336
ω_{12}	0.184	0.239	0.246	0.279	0.212	0.251	0.277
ω_{13}	0.164	0.106	0.194	0.049	0.272	0.074	0.367
a_{11}	0.078	0.077	0.070	0.079	0.027	0.033	0.048
a_{12}	0.012	0.014	0.012	0.017	0.007	0.008	0.038
a_{13}	0.200	0.202	0.192	0.211	0.061	0.071	0.077
a_{21}	0.012	0.028	0.020	0.031	0.033	0.034	0.042
a_{22}	0.005	0.009	0.009	0.010	0.010	0.014	0.014
a_{23}	0.100	0.088	0.068	0.106	0.071	0.095	0.102
a_{31}	0.150	0.152	0.148	0.151	0.014	0.016	0.019
a_{32}	0.029	0.029	0.030	0.030	0.004	0.005	0.010
a_{33}	0.120	0.113	0.113	0.113	0.032	0.035	0.037
b_{11}	0.743	0.747	0.637	0.743	0.186	0.201	0.276
b_{12}	0.031	0.018	0.113	0.024	0.205	0.193	0.308
b_{13}	−0.060	−0.064	0.021	−0.076	0.133	0.064	0.201
b_{21}	−0.020	0.041	0.046	0.108	0.216	0.120	0.317
b_{22}	0.851	0.792	0.752	0.726	0.221	0.206	0.312
b_{23}	0.053	−0.003	0.078	−0.096	0.179	0.148	0.285
b_{31}	−0.120	−0.196	0.003	−0.259	0.250	0.009	0.340
b_{32}	0.111	0.204	0.024	0.269	0.274	0.042	0.367
b_{33}	0.548	0.562	0.470	0.572	0.116	0.055	0.191
			Bias			Std.	
Average		0.025	0.045	0.048	0.122	0.075	0.177

Notes: The true parameter values are reported in the first column.
Case I imposes the matrix inequality constraints of Theorem 1.
Case II enforces Bollerslev's sufficient conditions. In case III no constraints are imposed.
For the latter case, there are about 20 out of the 1000 cases where negative
conditional means appear in the simulation/optimization. We disregard these cases.
Std. stands for the standard deviation. The last row reports the average bias and average standard deviation.

Root mean square errors (RMSE) for the out of sample forecasting are reported in Table 3. For one step ahead forecasting, the estimation imposing our matrix inequality constraints works best. For five step ahead forecasting, the estimation for the first two cases works equally well. For twenty step ahead forecasting, the estimated model based on Bollerslev's sufficient conditions displays the best performance, mainly because its QML estimates have smaller standard deviations. The estimation without enforcing any non-negativity constraints is the worst.

Table 3. RMSE for the out of sample forecasting.

Model	$k=1$			$k=5$			$k=20$		
	Case I	Case II	Case III	Case I	Case II	Case III	Case I	Case II	Case III
$\boldsymbol{\sigma}_{1,t}$	0.142	0.314	1.017	4.572	4.534	14.887	9.903	6.667	48.354
$\boldsymbol{\sigma}_{2,t}$	0.270	0.493	1.310	2.888	3.131	29.258	7.623	4.823	26.537
$\boldsymbol{\sigma}_{3,t}$	0.113	0.182	0.659	3.013	3.300	26.813	8.194	4.932	27.808

Notes: $k = 1, 5, 20$ are one, five and twenty days ahead forecasting, respectively.
Cases I, II and III are as in Table 2. To obtain the forecasts we use eq. (B.1) in Appendix B.

5 Extensions

As noted in the Introduction the constraints that all the parameters in multivariate extended GARCH models and MEM should be non-negative are only sufficient conditions and may be weakened in certain cases. In the previous Section we reviewed and generalized the non-negativity conditions derived in Conrad and Karanasos (2010). We proposed simple matrix inequalities which incorporate these conditions. In other words, we reformulated this set of constraints in a more (compact) intuitive visual form, that is we expressed them in a more efficient way as matrix inequalities. In the next Section we will examine various extensions such as asymmetric systems, higher shock orders and some new mixture formulations. Our methodology enables us to deal with these cases without difficulty.

5.1 Asymmetric Systems

In the asymmetric version of the N-dimensional system it will suffice to show that the matrix inequality constraints given in Theorem 1 are satisfied not only for $\mathbf{A}$ but for $\mathbf{A}+\mathbf{\Gamma}$ as well. In other words, it suffices to show that the constraints are satisfied in the two extreme cases: i) that is all the N *i.i.d* errors are non-negative for every t: $\mathbf{e}_t > 0$ for all t, and thus $\mathbf{A}_t = \mathbf{A}$ in eq. (2), since $\mathbf{s}_t = 0$ for all t, and ii) for every t $\mathbf{A}_t = \mathbf{A}+\mathbf{\Gamma}$, that is, $\mathbf{e}_t < 0$ for all t and, therefore, $\mathbf{s}_t = \mathbf{j}$. Then the following theorem holds.

Theorem 2 *Consider the N-dimensional vector SUE-AP model in eq. (2) and let Assumptions (A1) and (A2) in Appendix A be satisfied. Then, necessary and sufficient conditions for $\sigma_{it}^{\delta_i} > 0$, $i = 1, \dots, N$, for all t are given by:*
(a) $\boldsymbol{\mu} = adj[\mathbf{I}-\mathbf{B}]\boldsymbol{\omega} > \mathbf{0}$

$$(b)\ \begin{cases} \phi_1 \text{ is real, and } \phi_1 > 0, & (C1)\\ adj[\mathbf{I}\phi_1-\mathbf{B}]\mathbf{A} > \mathbf{0} \text{ and } adj[\mathbf{I}\phi_1-\mathbf{B}](\mathbf{A}+\mathbf{\Gamma}) > \mathbf{0}, & (C2^*)\\ \mathbf{B}^{k_a-1}\mathbf{A} \underset{(for\ k_a=1,\dots,\kappa_\alpha)}{\overset{(k_\alpha\le\kappa_{ij,\alpha})}{\succeq}} \mathbf{0} \text{ and } \mathbf{B}^{k_\gamma-1}(\mathbf{A}+\mathbf{\Gamma}) \underset{(for\ k_\gamma=1,\dots,\kappa_\gamma)}{\overset{(k_\gamma\le\kappa_{ij,\gamma})}{\succeq}} \mathbf{0} & (C3^*) \end{cases}$$

(the symbol $\overset{(k_\varsigma\le\kappa_{ij,\varsigma})}{\succeq}$ means that if $k_\varsigma > \kappa_{ij,\varsigma}$, $\varsigma = \alpha, \gamma$, then the ijth scalar inequality of the matrix inequality $\mathbf{\Psi}_{k_\varsigma} \geq \mathbf{0}$ should be disregarded)
(b^*)The $\kappa_{ij,\varsigma}$ and κ_ς in Condition (C3) can be obtained as follows: Let $\kappa_\varsigma = \max[\mathbf{K}^{(\varsigma)}]$, $\mathbf{K}^{(\varsigma)} = [\kappa_{ij,\varsigma}]$, and $\kappa_{ij,\varsigma}$ is the smallest integer greater than or equal to $\max\{0, \varphi_{ij,\varsigma}\}$ with $\mathbf{\Phi}^{(\varsigma)} = [\varphi_{ij,\varsigma}]$, given by*

$$\mathbf{\Phi}^{(\varsigma)} = [\log[\mathbf{H}^{(\varsigma,1)}] - \log[(N-1)\mathbf{H}_{\max}^{(\varsigma,n)}]][\log(|\phi_2|) - \log(|\phi_1|)]^{-1},$$

where $\mathbf{H}_{\max}^{(\varsigma,n)} = [\max_{2\le n\le N} \eta_{ij,\varsigma}^{(n)}]$ and $\mathbf{H}^{(\varsigma,n)} = [\eta_{ij,\varsigma}^{(n)}]$, $1 \le n \le N$, is given by

$$\mathbf{H}^{(\varsigma,n)} = abs\left[\frac{adj[\mathbf{I}\phi_n - \mathbf{B}]\mathbf{A}^{(\varsigma)}}{\sum_{j=1}^{N} j\beta_j \phi_n^{N-(j-1)}}\right],$$

with $\mathbf{A}^{(\varsigma)} = \mathbf{A}$ if $\varsigma = \alpha$ and $\mathbf{A}^{(\varsigma)} = \mathbf{A}+\mathbf{\Gamma}$ if $\varsigma = \gamma$.

Theorem 2 is similar to Theorem 1 with the only difference that we augment Conditions (C2) and (C3) in order to take into account the asymmetries. As with Theorem 1 in practice instead of checking Condition (C3*) one should check the matrix inequalities: $\mathbf{B}^{k_a-1}\mathbf{A} \geq \mathbf{0}$ and $\mathbf{B}^{k_\gamma-1}(\mathbf{A}+\mathbf{\Gamma}) \geq \mathbf{0}$ for all k_a and k_γ up to a large enough κ_a and κ_γ, i.e., $\kappa_a = \kappa_\gamma = N$. As we noted in Theorem 1 this simplification of Condition (C3*) is another important consequence of our matrix inequalities. Clearly, the latter conditions, for $k_a = k_\gamma = 1$, imply that $\mathbf{A}, (\mathbf{A}+\mathbf{\Gamma}) \geq \mathbf{0}$.

5.2 Models of Higher Order: $(1, q)$

In this Section we extend the order of the N-dimensional vector SUE-AP system in eq. (2) from $(1, 1)$ to $(1, q)$. That is we consider the multivariate process:

$$(\mathbf{I} - \mathbf{B}L)\boldsymbol{\sigma}_t = \boldsymbol{\omega} + \sum\nolimits_{l=1}^{q} L^l \mathbf{A}_{lt}\boldsymbol{\varepsilon}_t, \tag{9}$$

where $\mathbf{A}_{lt} = \mathbf{A}_l + \mathbf{\Gamma}_{lt}$, with $\mathbf{A}_l = [\alpha_{ij}^{(l)}]$, and $\mathbf{\Gamma}_{lt} = \mathbf{\Gamma}_l diag[\mathbf{s}_t]$, with $\mathbf{\Gamma}_l = [\gamma_{ij}^{(l)}]$ (recall that $\boldsymbol{\varepsilon}_t$ and $\mathbf{s}_t$ have been defined in eqs. (1) and (2), respectively). For the above process the following theorem holds.

Theorem 3 *Consider the N-dimensional vector SUE-AP $(1, q)$ model in eq. (9) and let Assumptions (A1)-(A2) in Appendix A be satisfied.*[5] *Then, the necessary and sufficient conditions for $\sigma_{it}^{\delta_i} > 0$, $i = 1, \ldots, N$, for all t, are as in Theorem 2 where we replace*
i) in Condition (C2), $\mathbf{A}$ and $\mathbf{\Gamma}$ by $\widetilde{\mathbf{A}} = \sum_{l=1}^{q} \mathbf{A}_l \phi_1^{q-l}$, and $\widetilde{\mathbf{\Gamma}} = \sum_{l=1}^{q} \mathbf{\Gamma}_l \phi_1^{q-l}$, respectively, and*
ii) in Condition (C3), $\mathbf{B}^{k_a - 1}\mathbf{A}$ by $\sum_{s=1}^{\min(q,k_\alpha)} \mathbf{B}^{k_\alpha - s}\mathbf{A}_s$, and $\mathbf{B}^{(k_\gamma - 1)}(\mathbf{A} + \mathbf{\Gamma})$ by $\sum_{s=1}^{\min(q,k_\gamma)} \mathbf{B}^{k_\gamma - s}(\mathbf{A}_s + \mathbf{\Gamma}_s)$*

(the proof is trivial). An analogous result (not reported) holds for the N-dimensional SUE-AP model 2 of order $(1, q)$. Clearly, if $q = 1$, then Theorem 3 becomes identical to Theorem 2.

In this Section we have presented results regarding the matrix inequality constraints for two important extensions of the SUE-P $(1, 1)$ model, namely, the asymmetric case and the higher order $(1, q)$ case. In the next Section we propose two new multivariate processes.

6 New Formulations

In this Section we examine two important developments. These are the asymmetric power mixture (APM) and the UE asymmetric power semi-exponential (APSE) model. The former specification can be used if some of the matrix inequality constraints in the SUE-AP model are violated. Once we have identified for which equation (power transformed conditional variable) the non-negativity conditions are not met, then they are easily removed by using the new mixture formulation.

The UE-APSE model allows not only for negative conditional spillovers but for negative shock (unconditional) spillover effects as well and, therefore, can be used as an alternative to the APM formulation.

6.1 Asymmetric Power Mixture Model

If a SUE-AP model is estimated and the non-negativity conditions are violated then what should be the next step? One solution to this problem is to use a multivariate exponential type of model, that is a system where we model instead of the power transformed conditional variables $(\sigma_{it}^{\delta_i})$ their logarithmic transformations, $ln(\sigma_{it})$. In what follows we will provide an alternative strategy which includes the exponential type of system as a special case. Before we do that, however, we will introduce some additional notation.

Define the N-dimensional vector $\boldsymbol{\sigma}_{M,t} = [\sigma_{M,it}]$, where

$$\sigma_{M,it} = \begin{cases} \sigma_{it}^{\delta_i} & \text{if} \quad i = 1, \ldots, d, \;\; d \geq 1 \\ \ln(\sigma_{it}) & \text{if} \quad i = d+1, \ldots, N, \;\; d \leq N-1. \end{cases}$$

with $d = 0, \ldots, N$. In other words the first d conditional variables in this N-dimensional mixture formulation are modelled as power specifications (as in Ding et al., 1993) whereas the rest are modelled as exponential processes (in the spirit of Nelson, 1991). The vector $\boldsymbol{\sigma}_{M,t}$ can be written as:

[5] with $\mathbf{A}L$ replaced by $\sum_{l=1}^{q} \mathbf{A}_l L^l$ and $(\mathbf{A} + \mathbf{\Gamma})L$ replaced by $\sum_{l=1}^{q}(\mathbf{A}_l + \mathbf{\Gamma}_l)L^l$.

$$\boldsymbol{\sigma}_{M,t} = \mathbf{I}_{[d]}\boldsymbol{\sigma}_t + \mathbf{I}_{(d)}\boldsymbol{\sigma}_{L,t}, \tag{10}$$

where $\boldsymbol{\sigma}_{L,t} = [\ln(\sigma_{it})]$ (recall that $\boldsymbol{\sigma}_t$ has been defined in eq. (1)), $\mathbf{I}_{[d]}$ is a diagonal matrix with ones in the first d diagonal elements and zeros elsewhere, and $\mathbf{I}_{(d)} = \mathbf{I} - \mathbf{I}_{[d]}$. Clearly, if $d = N$, then $\mathbf{I}_{[N]} = \mathbf{I}$, $\mathbf{I}_{(N)} = \mathbf{0}$, and $\boldsymbol{\sigma}_{M,t} = \boldsymbol{\sigma}_t$, whereas if $d = 0$, then $\mathbf{I}_{[0]} = \mathbf{0}$, $\mathbf{I}_{(0)} = \mathbf{I}$, and $\boldsymbol{\sigma}_{M,t} = \boldsymbol{\sigma}_{L,t}$.

Similarly, define the N-dimensional vector $\boldsymbol{\varepsilon}_{M,t} = [\varepsilon_{M,it}]$, where

$$\varepsilon_{M,it} = \begin{cases} |\varepsilon_{it}|^{\delta_i} & \text{if} \quad i = 1,\ldots,d,\ d \geq 1 \\ |e_{it}|^{\delta_i} & \text{if} \quad i = d+1,\ldots,N,\ d \leq N-1. \end{cases}$$

This vector can be expressed as:

$$\boldsymbol{\varepsilon}_{M,t} = \mathbf{I}_{[d]}\boldsymbol{\varepsilon}_t + \mathbf{I}_{(d)}\mathbf{z}_t \tag{11}$$

(recall that $\boldsymbol{\varepsilon}_t$ and $\mathbf{z}_t$ have been defined in eq. (1)). Clearly, if $d = N$, then $\boldsymbol{\varepsilon}_{M,t} = \boldsymbol{\varepsilon}_t$, whereas when $d = 0$, $\boldsymbol{\varepsilon}_{M,t} = \mathbf{z}_t$.

Then the N-dimensional vector SUE-APM model 1, that is, with the Glosten et al., 1993, type of asymmetry (we refer to the specification with the Ding et al., 1993 type of asymmetry as model 2; see Section B in the supplementary Appendix) consists of the following equations:

$$\begin{aligned} \sigma_{M,it} \quad = \quad & \omega_i + \sum\nolimits_{j=1}^{d} (\alpha_{ij} + \gamma_{ij} s_{j,t-1}) \left|\varepsilon_{j,t-1}\right|^{\delta_j} + \sum\nolimits_{j=d+1}^{N} (\alpha_{ij} + \gamma_{ij} s_{j,t-1}) \left|e_{j,t-1}\right|^{\delta_j} \\ & + \sum\nolimits_{j=1}^{d} \beta_{ij}\sigma_{j,t-1}^{\delta_j} + \sum\nolimits_{j=d+1}^{N} \beta_{ij} \ln(\sigma_{j,t-1}),\ i = 1,\ldots,N. \end{aligned}$$

The system in a matrix form can be written as

$$(\mathbf{I} - \mathbf{B}L)\boldsymbol{\sigma}_{M,t} = \boldsymbol{\omega} + L\mathbf{A}_t\boldsymbol{\varepsilon}_{M,t}, \tag{12}$$

where $\mathbf{B}$, $\boldsymbol{\omega}$ and $\mathbf{A}_t$ are as in eq. (2). When $d = N$, the APM model becomes identical to the AP model in eq. (2), whereas if $d = 0$, it reduces to the multivariate extension of the exponential specification of Nelson (1991):

$$(\mathbf{I} - \mathbf{B}L)\boldsymbol{\sigma}_{L,t} = \boldsymbol{\omega} + L\mathbf{A}_t\mathbf{z}_t. \tag{13}$$

We will refer to this model as the UE asymmetric power exponential (UE-APE). This multivariate process in the context of MEM has been used in Taylor and Xu (2017). An applied researcher can also use a number of alternative multivariate APM specifications, i.e., replacing $\mathbf{z}_t$ in eq. (11) by either $\boldsymbol{\varepsilon}_t$ or $\boldsymbol{\varepsilon}_{L,t} = [\ln(\varepsilon_{it}^2)]$ (see, for details, Section B in the supplementary Appendix). For all these models the non-negativity constraints are identical and will be given in the following theorem. But first, recall that the N order matrix $\mathbf{J}_{[d]}$ indicates a binary matrix with ones in its first d rows and zeros elsewhere; when $d = 0$, then $\mathbf{J}_{[0]} = \mathbf{0}$, whereas when $d = N$, $\mathbf{J}_{[d]}$ becomes the unit matrix $\mathbf{J}$.

Theorem 4 *Consider the N-dimensional vector SUE-APM model in eq. (12) and let Assumptions (A1)-(A2) in Appendix A be satisfied. Then, necessary and sufficient conditions for $\sigma_{it}^{\delta_i} > 0$, $i = 1,\ldots,d$, for all t in Theorem 2 become*

(a) $\boldsymbol{\mu} = adj[\mathbf{I} - \mathbf{B}]\boldsymbol{\omega} \odot \mathbf{J}_{[\mathbf{d}]} \overset{d}{\succ} \mathbf{0}$

(b)
$$\begin{cases} \phi_1 \text{ is real, and } \phi_1 > 0, & \text{(C1)} \\ adj[\mathbf{I}\phi_1 - \mathbf{B}]\mathbf{A} \odot \mathbf{J}_{[\mathbf{d}]} \overset{d}{\succ} \mathbf{0} \text{ and } adj[\mathbf{I}\phi_1 - \mathbf{B}](\mathbf{A} + \boldsymbol{\Gamma}) \odot \mathbf{J}_{[\mathbf{d}]} \overset{d}{\succ} \mathbf{0}, & \text{(C2**)} \\ \underset{(for\ k_a = 1,\ldots,\kappa_\alpha)}{(\mathbf{B}^{k_a - 1}\mathbf{A}) \odot \mathbf{J}_{[\mathbf{d}]}} \overset{(d,\ k_\alpha \leq \kappa_{ij,\alpha})}{\succeq} \mathbf{0}, \text{ and} & \\ & \text{(C3**)} \\ \underset{(for\ k_\gamma = 1,\ldots,\kappa_\gamma)}{[\mathbf{B}^{k_\gamma - 1}(\mathbf{A} + \boldsymbol{\Gamma})] \odot \mathbf{J}_{[\mathbf{d}]}} \overset{(d,\ k_\gamma \leq \kappa_{ij,\gamma})}{\succeq} \mathbf{0} & \end{cases}$$

(the symbol $\overset{d}{\succ}$ means that we check only the inequalities in the first d rows of the matrix and we disregard the other inequalities) where $\kappa_\varsigma = \max[\mathbf{K}^{(\varsigma)} \odot \mathbf{J}_{[\mathbf{d}]}]$ and $\mathbf{K}^{(\varsigma)}$, $\varsigma = \alpha, \gamma$, are as in Theorem 2.

The non-negativity constraints (matrix inequalities) in the above theorem are similar to (actually exactly the same as) those in Theorem 2 with the only difference that now for each matrix inequality we have only $d \times N$ scalar inequalities and not N^2, that is, we only have to check the scalar inequalities in the first d rows which are linked to the d power transformed conditional variables. In other words, if $d = N$, then the above Theorem becomes identical to Theorem 2, whereas when $d = 0$, it becomes redundant, since in the UE-APE model no constraints are needed.

6.2 Asymmetric Power Semi Exponential Model

Here we will present the new N-dimensional UE-APSE process. This formulation, with the Glosten et al. (1993) type of asymmetry (model 1), is given by

$$\sigma_{it}^{\delta_i} = \omega_i + \sum_{j=1}^{N}\sum_{l=1}^{N} \theta_{ij} \exp[(\alpha_{jl} + \gamma_{jl} s_{j,t-1}) |e_{j,t-1}|^{\delta j}] + \sum_{j=1}^{N} \beta_{ij} \sigma_{j,t-1}^{\delta_j},\ i = 1, \dots, N.$$

The above process can be written in a matrix form as

$$(\mathbf{I} - \mathbf{B}L)\boldsymbol{\sigma}_t = \boldsymbol{\omega} + \boldsymbol{\Theta} L \mathbf{e}^{\mathbf{A}_t \mathbf{z}_t}, \tag{14}$$

where $\boldsymbol{\Theta} = [\theta_{ij}]$ and $e^{\mathbf{x}}$ means that we take the exponential of each element of the $\mathbf{x}$ vector. Thus, since we take the exponential of each element $\alpha_{ij} + \gamma_{ij} s_{j,t-1}$ this model allows matrices $\mathbf{A}$ and $\boldsymbol{\Gamma}$ to take any negative values. However, the $\mathbf{B}$ and $\boldsymbol{\Theta}$ matrices must satisfy the matrix inequality constraints of Theorem 1. Therefore, the following theorem holds.

Theorem 5 *Consider the N-dimensional UE-APSE model in eq. (14) and let Assumptions (A1)-(A2) (with $\mathbf{A}$ replaced by $\boldsymbol{\Theta}$) in Appendix A be satisfied. Then, necessary and sufficient conditions for $\sigma_{it}^{\delta_i} > 0$, $i = 1, \dots, N$, for all t, are as in Theorem 1 with $\mathbf{A}$ replaced by $\boldsymbol{\Theta}$.*

The next table presents a summary of the various multivariate models and theoretical results.

Table 4. Various GARCH/HEAVY models and MEM.

Models↓	Formulations	Notation	Non-negativity Constraints
SUE-P	$(\mathbf{I} - \mathbf{B}L)\boldsymbol{\sigma}_t = \boldsymbol{\omega} + L\mathbf{A}\boldsymbol{\varepsilon}_t$ [Eq. (3)]	$\boldsymbol{\sigma}_t = [\sigma_{it}^{\delta_i}]$, $\boldsymbol{\varepsilon}_t = \mathbf{Z}_t\boldsymbol{\sigma}_t$ $\mathbf{Z}_t = diag[\mathbf{z}_t]$, $\mathbf{z}_t = [\lvert e_{it}\rvert^{\delta_i}]$ [Eq. 1]	Theorem 1
SUE-AP	$(\mathbf{I} - \mathbf{B}L)\boldsymbol{\sigma}_t = \boldsymbol{\omega} + L\mathbf{A}_t\boldsymbol{\varepsilon}_t$ [Eq. (2)]	$\mathbf{A}_t = \mathbf{A} + \boldsymbol{\Gamma}_t$, $\boldsymbol{\Gamma}_t = \boldsymbol{\Gamma} diag[\mathbf{s}_t]$	Theorem 2
SUE-APM	$(\mathbf{I} - \mathbf{B}L)\boldsymbol{\sigma}_{M,t} = \boldsymbol{\omega} + L\mathbf{A}_t\boldsymbol{\epsilon}_{M,t}$ [Eq. (12)]	$\boldsymbol{\sigma}_{M,t} = \mathbf{I}_{[d]}\boldsymbol{\sigma}_t + \mathbf{I}_{(d)}\boldsymbol{\sigma}_{L,t}$ [Eq. (10)] $\boldsymbol{\varepsilon}_{M,t} = \mathbf{I}_{[d]}\boldsymbol{\varepsilon}_t + \mathbf{I}_{(d)}\mathbf{z}_t$ [Eq. (11)]	Theorem 4
UE-APSE	$(\mathbf{I} - \mathbf{B}L)\boldsymbol{\sigma}_t = \boldsymbol{\omega} + \boldsymbol{\Theta} L \mathbf{e}^{\mathbf{A}_t\mathbf{z}_t}$, [Eq.(14)]	$\boldsymbol{\Theta} = [\boldsymbol{\theta}_{ij}]$	Theorem 5
UE-APE	$(\mathbf{I} - \mathbf{B}L)\boldsymbol{\sigma}_{L,t} = \boldsymbol{\omega} + L\mathbf{A}_t\mathbf{z}_t$ [Eq.(13)]	$\boldsymbol{\sigma}_{L,t} = [\ln(\sigma_{it})]$	No constraints

7 Empirical Results

In this Section we estimate trivariate and four variate SUE MEM systems (that is symmetric models with $\delta_i = 1$ for all i; see also Cipollini, et al. 2013):

$$(\mathbf{I} - \mathbf{B}L)\widetilde{\boldsymbol{\sigma}}_t = \boldsymbol{\omega} + \mathbf{A}L\widetilde{\boldsymbol{\varepsilon}}_t,$$

where we recall that $\widetilde{\boldsymbol{\sigma}}_t = [\sigma_{it}]$ and $\widetilde{\boldsymbol{\varepsilon}}_t = \mathbf{E}_t\widetilde{\boldsymbol{\sigma}}_t$ with $\mathbf{E}_t = diag[\mathbf{e}_t]$, $\mathbf{e}_t = [e_{it}]$ (see Section 2.2). Notice that $\widetilde{\boldsymbol{\varepsilon}}_t$, is the vector, which contains the observed series. Since we now use the MEM, we assume that the stochastic vector $\mathbf{e}_t > \mathbf{0}$ (and, hence, $\widetilde{\boldsymbol{\varepsilon}}_t > \mathbf{0}$) is $i.i.d$ with unit vector $\mathbf{j}$ as an expectation, positive definite correlation matrix $\mathbf{R} = [\rho_{ij}]$ with $\rho_{ii} = 1$, and covariance matrix $\mathbf{Q} = [q_{ij}] = diag[\mathbf{q}^{\wedge 1/2}]\mathbf{R}diag[\mathbf{q}^{\wedge 1/2}]$ with $\mathbf{q} = [q_{ii}]$. So now, $\widetilde{\boldsymbol{\sigma}}_t = \mathbb{E}(\widetilde{\boldsymbol{\varepsilon}}_t \,|\mathcal{F}_{t-1})$ (see also Section 2.2). The elements of $\widetilde{\boldsymbol{\varepsilon}}_t$ could be the intraday trading duration, volume and volatility, or three different measures of volatility (i.e., high-low range volatility, absolute return and realized volatility) for an individual asset, or the volatility proxy (i.e., high-low range volatility) for several financial markets.

Following Taylor and Xu (2017) we use the multivariate log-normal distribution for the innovation vector $\mathbf{e}_t$, which is a random vector defined in $[\mathbf{0}, +\infty)$, that is $\mathbf{e}_t \sim \ln N(\mathbf{j}, \mathbf{Q})$. The log likelihood function, based on $\widetilde{\boldsymbol{\varepsilon}}_t = \mathbf{E}_t\widetilde{\boldsymbol{\sigma}}_t$, is given by

$$l(\boldsymbol{\theta}) = \sum\nolimits_{t=1}^{T} \ln f(\widetilde{\boldsymbol{\varepsilon}}_t|\boldsymbol{\theta}),$$

where

$$\ln f(\widetilde{\boldsymbol{\varepsilon}}_t|\boldsymbol{\theta}) = -\frac{N}{2}\ln(2\pi) - \frac{1}{2}\ln|\mathbf{Q}| - \sum\nolimits_{j=1}^{N} \ln \varepsilon_{jt} - \frac{1}{2}(\ln\widetilde{\boldsymbol{\varepsilon}}_t - \ln\widetilde{\boldsymbol{\sigma}}_t - \boldsymbol{\xi})'\mathbf{Q}^{-1}(\ln\widetilde{\boldsymbol{\varepsilon}}_t - \ln\widetilde{\boldsymbol{\sigma}}_t - \boldsymbol{\xi}), \quad (15)$$

with $\boldsymbol{\xi} = -1/2\mathbf{q}$. These assume that $\boldsymbol{\theta} = [\boldsymbol{\beta}', \boldsymbol{\rho}]$, where $\boldsymbol{\beta}$ contains the parameters in $\widetilde{\boldsymbol{\sigma}}_t$, and $\boldsymbol{\rho} = vech(\mathbf{Q})$. Here, the vech operator stacks the lower triangular elements of the N order symmetric $\mathbf{Q}$ matrix into the $(N \times (N+1)/2)$ $\boldsymbol{\rho}$ vector.

The model can be estimated consistently by QML estimation.[6] In what follows we estimate three (two trivariate and one four-variate) SUE MEM, based on data and model availability. The first example is a trivariate system of intra day trading duration, stock volume and volatility. We use the same dataset as in Manganelli (2005), who estimated an equation-by-equation specification of this model, but we estimate a trivariate system.[7] The second model employs three volatility indicators (high-low range volatility, absolute return and realized volatility), which was proposed by Engle and Gallo (2006), and also estimated by Cipollini et al. (2013).[8] The third example is by Cipollini et al. (2010), who estimated a four-variate process by using daily high-low range data in four EU stock markets (UK, France, Germany, Switzerland). We use the latest data (from 01/01/2003 to 31/12/2014). As noted earlier all MEM are estimated by employing the QML estimation strategy initially proposed by Taylor and Xu (2017). The estimation results are reported in the Tables below. We restrict the $\mathbf{A}$ matrix to be non-negative. That is, combining restrictions on the parameter space ($\mathbf{A} \geq \mathbf{0}$) with the sample data, we are using the restricted QML estimation.

Duration, Volume and Volatility

Manganelli (2005) used a diagonal $\mathbf{B}$ matrix and an unrestricted $\mathbf{A}$ matrix. It should be noted that Manganelli's results violate the non-negativity conditions in Theorem 1 (in Section I of the supplementary Appendix we give examples where well known papers report estimated parameters that violate these conditions). Caution should be used when estimating a model where the $\mathbf{A}$ matrix is unrestricted. In our estimation all the parameters of the $\mathbf{A}$ matrix that turned out to be negative were set equal to zero (see Table 5 below). For example, in line with Manganelli, who reports negative α_{13} and α_{31}, in almost all cases these two parameters took negative values and, therefore, were set equal to zero. Similarly, all the parameters of the $\mathbf{B}$ matrix that turned out to be negative and violated the matrix inequality constraints were set equal to zero.

In particular, the conditional mean of trading duration affects that of stock volatility positively in four out of the five cases (β_{31} is positive and significant). This result is in line with the predictions of

[6] An alternative estimation method was proposed by Cipollini et al. (2013). They bypassed the specification conditional distribution of the errors and made use of only the first two conditional moments of the errors by using an efficient generalized methods of moments (GMM) estimation method. By a simulation study Taylor and Xu (2017) showed that both the QML and GMM estimation techniques are consistent and that the efficient loss of the QML estimation compared with the GMM estimation due to misspecification of the error distribution is trivial.

[7] See Section H in the supplementary Appendix or Subsection 4.1 in Manganelli (2005) for a concise description of how the data are prepared for use in his and our paper.

[8] For the description of the data set see Section H in the supplementary Appendix or Cipollini et al. (2013).

Diamond and Verrecchia (1987). For CP and DLP the conditional mean of stock volume has a negative impact on that of volatility (β_{32} is negative and significant), which is in line with the theory by Wang (2007). According to Wang foreign purchases tend to lower volatility by increasing the investor base in emerging markets, since the broadening of the investor base improves the accuracy of market information and stabilizes stock prices (see also Karanasos and Kartsaklas, 2009). Clearly, the negative estimated parameter β_{32} would have been ruled out by the sufficient Bollerslev conditions. It also appears that the conditional mean of trading duration is independent of changes in the other two conditional means as in all cases (except one) the β_{12} and β_{13} parameters are either set equal to zero or are insignificant. Similarly, the conditional mean of stock volume is independent of changes in the conditional means of the other two variables as in four out of the five cases the β_{21} and β_{23} parameters are set equal to zero. Only for the COX case does the conditional mean of volatility affect that of stock volume negatively. This result is in alignment with the work of Li and Wu (2006). Clearly, the negative value of β_{23} would have been ruled out by the sufficient Bollerslev conditions. It is easy to check that the matrix inequality constraints of Theorem 1 are satisfied for the given parameter combination (see Table G1 in Section G of the supplementary Appendix).

Table 5. Trivariate SUE MEM (1, 1) of intra day trading duration, stock volume and volatility.

	AVT			COX			CP			DLP			GAP		
A	**0.141**	-	-	**0.132**	-	-	**0.044**	-	-	**0.090**	-	-	**0.101**	-	-
	(14.6)			(21.49)			(10.79)			(16.06)			(13.60)		
	0.004	0.045	-	-	**0.004**	**0.006**	-	**0.029**	**0.003**	-	**0.050**	-	-	**0.080**	-
	(0.55)	(1.29)			(9.13)	(3.72)		(2.04)	(2.03)		(4.12)			(4.76)	
	-	0.063	**0.288**	-	**0.041**	**0.314**	**0.151**	**0.424**	**0.768**	-	**0.057**	**0.267**	-	**0.527**	**0.748**
		(1.03)	(16.4)		(7.29)	(50.33)	(29.60)	(21.71)	(107.8)		(5.90)	(34.75)		(13.72)	(50.21)
B	**0.873**	-	-	**0.873**	0.001	-	**0.950**	0.001	-	**0.912**	**0.004**	0.000	**0.921**	-	-
	(90.2)			(148.3)	(0.18)		(176.3)	(0.69)		(169.1)	(3.14)	(0.72)	(149.6)		
	-	**0.775**	-	**0.001**	**0.987**	**-0.004**	-	**0.873**	-	-	**0.858**	-	-	**0.429**	-
		(3.12)		(3.04)	(587.8)	(5.64)		(11.97)			(20.8)			(2.98)	
	0.017	-0.223	**0.826**	0.002	0.003	**0.686**	**0.033**	**-0.040**	**0.232**	**0.024**	**-0.093**	**0.826**	**0.529**	-	**0.252**
	(2.21)	(0.75)	(81.9)	(0.45)	(0.04)	(110.1)	(2.83)	(3.59)	(32.64)	(7.82)	(3.66)	(162.4)	(17.48)		(16.90)

Notes: We use the same data set as Manganelli (2005). Bollerslev-Wooldridge robust t-statistics in parentheses.
Variables significant at the 5 percent confidence level formatted in bold. 1st row: duration; 2nd row: volume; 3rd row: volatility.
The ω_i and the q_{ij}, $i, j = 1, \ldots, N$ are not reported but they are available upon request.

Table 6. Trivariate SUE MEM(1, 1) of daily high-low range volatility, absolute return and realized volatility.

	DJ30			S&P500			NASA			FTSE			DAX		
A	**0.058**	0.013	**0.193**	0.083	0.012	**0.137**	0.088	**0.020**	0.033	**0.086**	0.014	**0.100**	**0.150**	**0.019**	**0.095**
	(2.65)	(1.77)	(5.72)	(1.55)	(0.69)	(4.87)	(1.75)	(2.02)	(1.20)	(3.95)	(1.59)	(2.61)	(10.21)	(3.35)	(4.04)
	0.025	0.010	0.086	0.006	0.023	0.082	0.024	0.051	0.004	0.009	0.011	0.094	0.011	0.008	0.127
	(0.50)	(0.99)	(1.30)	(0.01)	(0.01)	(1.00)	(0.50)	(0.01)	(0.04)	(0.01)	(0.01)	(1.77)	(1.55)	(0.01)	(1.87)
	0.079	0.016	**0.269**	**0.060**	**0.027**	**0.280**	0.009	**0.029**	**0.337**	**0.068**	**0.029**	**0.204**	**0.099**	**0.038**	**0.216**
	(2.52)	(1.74)	(6.25)	(2.65)	(2.70)	(7.43)	(0.75)	(3.92)	(8.88)	(3.54)	(3.33)	(5.48)	(4.41)	(5.33)	(7.58)
B	**0.881**	**0.063**	**-0.240**	**0.862**	**0.045**	**-0.170**	**0.907**	0.010	**-0.059**	**0.806**	0.036	-0.065	**0.876**	**0.109**	**-0.281**
	(18.72)	(2.07)	(9.89)	(15.01)	(3.24)	(3.06)	(16.85)	(0.44)	(2.93)	(5.63)	(1.09)	(0.34)	(26.83)	(4.58)	(4.92)
	-0.062	**0.986**	-0.046	-0.011	**0.955**	-0.044	0.001	**0.937**	0.013	-0.015	**0.910**	0.019	-0.010	**0.958**	-0.087
	(0.52)	(10.68)	(0.31)	(0.04)	(22.50)	(0.14)	(0.02)	(45.39)	(0.13)	(0.69)	(15.05)	(0.13)	(1.81)	(13.57)	(0.53)
	-0.108	0.099	**0.599**	**-0.097**	0.098	**0.571**	-0.032	**0.067**	**0.558**	-0.046	**0.098**	**0.603**	-0.058	**0.164**	**0.491**
	(2.13)	(0.01)	(8.57)	(1.97)	(1.51)	(5.64)	(1.29)	(2.13)	(9.97)	(0.50)	(3.25)	(4.08)	(1.12)	(4.90)	(7.08)

Notes: We use the same data set as Cipollini et al. (2013). Bollerslev-Wooldridge robust t-statistics in parentheses.
Variables significant at the 5 percent confidence level formatted in bold.
1st row: high-low range volatility; 2nd row: absolute return; 3rd row: realized volatility.
The ω_i and the q_{ij}, $i, j = 1, \ldots, N$ are not reported but they are available upon request.

Three Volatility Measures

We also find significant dynamic interactions among the three different volatility measurements (daily high-low range volatility, absolute return, and realized volatility), which is consistent with Cipollini et al. (2013); see Table 6 above. However in the aforementioned paper, the estimated **A** matrix has negative elements and, therefore, the non-negativity conditions are violated (see Section I in the supplementary Appendix). As a result, negative values may be observed if their estimated model is used to forecast the three measures of volatility. Nevertheless, the results in Cipollini et al. (2013) indicate that there might exist negative interactions between the three conditional means. If we restrict the **B** matrix to be non-negative (as in the BEKK formulation used in Noureldin et al., 2012), some of the most important dynamic interconnections between the three volatility indicators may be lost and as a result the forecasts of these volatility measurements may not be as accurate as they should be. Therefore we allow some of the elements in the **B** matrix to be negative while at the same time we make sure that the matrix inequality constraints of Theorem 1 are preserved. In other words we estimate a less restricted model.

Indeed, our results show that β_{13} and β_{31} can be both negative (see for example the DJ30 and S&P 500 cases), while the non-negativity conditions are satisfied (see Table G2 in Section G of the supplementary Appendix). Cippolini et al. (2013) view the high-low range volatility as a proxy for jumps in the realized volatility. Our finding, that these jumps have a negative impact on the realized volatility ($\beta_{31} < 0$), is consistent with that in Andersen et al. (2007). This finding is also interesting, since it shows that negative conditional spillover effects in both directions are permitted by the matrix inequality constraints. In sharp contrast, such negative bidirectional feedback is prohibited in a bivariate restricted (or even semi unrestricted) extended system (see Conrad and Karanasos, 2010). Similarly to the DJ30 and S&P 500 indices in the other three cases the parameters β_{13} and β_{31} take negative values. Interestingly, in all five datasets the conditional means of the absolute returns are independent of changes in the other two conditional means. Finally, the positive and significant β_{32} parameter implies that in three out of the five indices the conditional mean of the absolute returns has a positive impact on that of the realized volatility. This result is in line with the finding of Forsberg and Ghysels (2007), that absolute return is the most favorable regressor for predicting realized volatility.

High-low Range Volatility in Four Equity Markets

Our results regarding the links between the high-low volatilities of the four European equity markets are presented in Table 7 below. Interestingly, seven out of the twelve off-diagonal elements of **A** are positive and significant. For example, the German and UK volatilities affect the conditional mean of Swiss volatility (see the fourth row), while the conditional mean of the French volatility (in the first row) is affected by all three volatilities. Most importantly, in the **B** matrix eight out of the twelve cross effects elements are negative (three of which significant) and yet the non-negativity conditions (matrix inequality constraints) of Theorem 1 are satisfied (see Table G3 in Section G of the supplementary Appendix). Interestingly, in the equations for France and UK (first and third rows) all six off-diagonal parameters are negative.

Most importantly, as pointed out by Cipollini and Gallo (2010), in the semi-unrestricted model, since it allows for negative conditional spillovers, the speed of absorption of a shock can be higher than in the restricted specification. For example, an increase in the UK high-low range at time $t-1$ will increase ($\alpha_{33} > 0$) its conditional mean at time t, which will be further increased at time $t+1$ ($\beta_{33} > 0$). However, the initial increase in the UK range will also boost the conditional range of Switzerland upwards ($\alpha_{43} > 0$), which will decrease the UK one at time $t+1$ ($\beta_{34} < 0$). The former effect will partially offset the latter. Cipollini and Galo (2010) also report negative values in their estimated **B** matrix but the matrix inequality constraints are violated (see Section I in the supplementary Appendix).

Table 7. Four variate SUE MEM $(1,1)$ of daily high-low range volatility in four Euro equity markets.

A	**0.080**	**0.042**	**0.040**	**0.033**
	(6.83)	(2.61)	(3.27)	(2.22)
	-	**0.161**	0.002	**0.028**
		(11.92)	(0.25)	(2.14)
	0.024	0.020	**0.086**	**0.049**
	(1.69)	(1.42)	(5.59)	(3.29)
	0.013	**0.030**	**0.024**	**0.108**
	(1.02)	(2.15)	(2.21)	(8.36)
B	**0.891**	**-0.050**	-0.012	-0.038
	(46.41)	(2.27)	(0.60)	(1.67)
	-	**0.804**	-	-
		(46.17)		
	-0.037	-0.029	**0.923**	**-0.049**
	(1.65)	(1.44)	(34.79)	(2.08)
	-0.044	-0.028	0.024	**0.853**
	(2.09)	(1.41)	(1.23)	(37.15)

Notes: We use the same data set as Cipollini and Gallo (2010). Bollerslev-Wooldridge robust t-statistics in parentheses. Variables significant at the 5 percent confidence level formatted in bold. 1st, 2nd, 3rd and 4th rows: FR, GE, UK and SW, respectively. The ω_i and the q_{ij}, $i,j = 1,\dots,N$ are not reported but they are available upon request.

Table 8. Trivariate SUE-AP MEM $(1,1)$ Model.

A	0.002	**0.147**	0.003
	(0.05)	(1.59)	(0.03)
	0.002	**0.114**	**0.054**
	(0.01)	(2.65)	(2.61)
	-	**0.123**	0.002
		(2.47)	(0.01)
B	**0.765**	0.003	0.003
	(30.53)	(0.01)	(0.01)
	0.002	**0.696**	-
	(0.05)	(3.45)	
	-	0.001	**0.741**
		(0.01)	(4.25)
Γ	0.011	**0.239**	0.002
	(0.24)	(2.06)	(0.02)
	0.127	**0.108**	0.001
	(2.56)	(2.22)	(0.03)
	0.095	0.109	0.030
	(4.06)	(1.48)	(0.51)
	Returns	Realized Volatility	GK Volatility
δ	**1.70**	**1.40**	**1.37**
	(7.05)	(17.00)	(13.64)

Notes: Bollerslev-Wooldridge robust t-statistics in parentheses. Variables significant at the 5 percent confidence level formatted in bold. 1st, 2nd, and 3rd rows: Returns, Real. Vol. and GK Vol., respectively. The ω_i and the q_{ij}, $i,j = 1,\dots,N$ are not reported but they are available upon request.

7.1 The SUE-AP Model

In this Section we estimate a trivariate SUE-AP system, as given by eq. (2), using daily data for the S&P 500 stock index from 03/01/2000 to 01/03/2013. The 3-dimensional process can be estimated either as a multivariate GARCH specification or as a MEM. In the first case, we model the power transformed conditional variances of the three variables, that is stock returns, SSR realized volatility and SSR Garman Klass (GK) volatility, using a multivariate normal distribution. For the MEM we model the power transformed conditional means of the three squared variables, that is squared returns, realized volatility and GK volatility, using a multivariate log-normal distribution. In what follows we will use the MEM (and QML estimation, see eq. (15)). The first equation in Table 8 is for the returns, and the other two for SSR realized and GK volatilities, respectively.

There are significant interactions between the three conditional means. The most dominant variable is the power transformed ($\delta = 1.40$) realized volatility since it has a significant impact on the power transformed ($\delta = 1.70$, 1.37) conditional means of the other two variables; see the second column of the $\mathbf{A}$ matrix. The stock return series is also an influential variable which affects the two volatilities. In particular, since the parameters γ_{21} and γ_{31} are significant, we find asymmetric shock spillovers from the power transformed ($\delta = 1.70$) stock returns to the power transformed (the two δs are 1.40 and 1.37) conditional means of the two volatilities. In other words, it is only for the negative returns that such cross effects are significant. The GK volatility is the less forcible of the three variables. These results are consistent with those presented in Karanasos et al. (2017). Interestingly, only the diagonal elements of $\mathbf{B}$ are significant, in other words there is no evidence of conditional spillovers either positive or negative.

7.2 UE-PSE Specification

Since the SUE model in eq. (2) restricts the elements of the $\mathbf{A}$ matrix to be non-negative we also estimate for the AVT data from the first dataset the following version of the UE-PSE model in eq. (14) (which allows the elements of the $\mathbf{A}$ matrix to take either positive or negative values):

$$(\mathbf{I} - \mathbf{B}L)\widetilde{\boldsymbol{\sigma}}_t = \boldsymbol{\omega} + \boldsymbol{\Theta}\mathbf{L}e^{\mathbf{A}\mathbf{z}_t^*},$$

where $\mathbf{z}_t^* = [|e_{1t}|, |e_{2t}|, |e_{3t}|]'$. The estimated (using QML estimation, see eq. (15)) $\mathbf{B}$ and $\boldsymbol{\Theta}$ matrices are now restricted to be diagonals (and therefore non-negative).

Our results are consistent with those in Table 5 (see the AVT part) where most of the off-diagonal elements of $\mathbf{B}$ were either restricted to zero or took negative values. Most importantly, there are negative bidirectional shock spillovers between duration and volatility (α_{13} and α_{31} are negative and significant), a finding that confirms many microstructure predictions. In particular, Easley and O'Hara (1992) have argued that times of high activity (short durations) are associated with a larger fraction of informed traders in the market, which leads to a quick adjustment in the prices and, hence, increases price volatility ($\alpha_{31} < 0$). Similarly, the negative value of α_{13}, shows that short durations follow large absolute price changes, and it is consistent with the findings of Engle (2000) and Manganelli (2005). This result implicitly suggest that large absolute quote changes indicate a risk of informed trading, which may make some liquidity traders leave the market or slow down their trading activity in order to avoid adverse selection, as predicted by the microstructure theory (see, for example, Easley and O'Hara, 1987, and Admati and Pfleiderer, 1988).

There is also a negative impact from volatility to volume (captured by the α_{23} parameter) as well. Interestingly, the three corresponding parameters in Table 5 (α_{13}, α_{31}, and α_{23}) were set to zero. The aforementioned negative bidirectional shock spillovers are in line with the ones in Manganelli (2005), who also reported negative α_{13} and α_{31} parameters. However, since he did not employ the semi-exponential specification, these negative values violate the matrix inequality constraint: $\mathbf{A} \geq \mathbf{0}$.

Table 9. Trivariate UE-PSE MEM.

AVT data from the first dataset			
A	**0.704**	-0.049	**-0.119**
	(29.57)	(0.83)	(2.80)
	0.010	**0.259**	**-0.110**
	(0.13)	(29.15)	(2.52)
	-0.164	**0.057**	**0.723**
	(3.24)	(2.11)	(35.56)
B	**0.770**	-	-
	(48.57)		
	-	**0.859**	-
		(28.76)	
	-	-	**0.736**
			(56.31)
$diag(\mathbf{\Theta})$	**0.196**	**0.093**	**0.299**
	(10.98)	(5.43)	(18.55)

Notes are as in Table 5.

7.3 Mixture Formulation

Having estimated two trivariate symmetric models (the SUE MEM and the UE-PSE one), using the AVT data from the first dataset, in this Section we turn our attention, for comparison purposes, to the trivariate mixture MEM formulation:

$$(\mathbf{I} - \mathbf{B}L)\boldsymbol{\sigma}_{M,t} = \boldsymbol{\omega} + \mathbf{A}L\boldsymbol{\varepsilon}_{M,t},$$

(see eq. (12)), where now $\boldsymbol{\sigma}_{M,t} = [\ln(\sigma_{1t}), \sigma_{2t}, \ln(\sigma_{3t}))]'$ and $\boldsymbol{\varepsilon}_{M,t} = [|e_{1t}|, \varepsilon_{2t}, |e_{3t}|]'$. In other words, we use the conditional mean of volume, whereas for duration and volatility we model the logarithms of their conditional means. The estimated results (using QML estimation, see eq. (15)) are presented in Table 10 below. Interestingly, there are significant negative spillovers from volume to duration (unconditional) and volatility (conditional), and conditional spillovers from volatility to duration (that is, α_{12}, β_{32} and β_{13}, respectively are negative and significant). This empirical exercise illustrates the importance of the mixture formulation. All these three significant negative impacts were either not permitted, or were insignificant, in the estimation of the SUE MEM (see the AVT part of Table 5).

Notice that for this mixture formulation the non-negativity conditions of only the second equation should be checked. Therefore the matrix inequality constraints of Theorem 4 are satisfied (see the numbers in italic in the bottom part of the Table).

Finally, comparing the results in the last two Tables, it seems that the negative α_{13} parameter in Table 9 has been regained by the corresponding negative β parameter in Table 10 (as predicted by the microstructure theory), whereas α_{32} is significant in both cases. In the last table β_{32} is also significant and negative, confirming the theoretical prediction of Wang (2007).

Table 10. Trivariate SUE-PM MEM.

AVT data from the first dataset			
A	**0.084**	**-0.010**	-0.005
	(3.26)	(2.83)	(0.15)
	-	**0.025**	-
		(7.38)	
	-0.005	**0.016**	**0.296**
	(0.18)	(6.48)	(7.64)
B	**0.915**	-	**-0.112**
	(46.40)		(18.09)
	-	**0.899**	-
		(54.54)	
	-	**-0.100**	**0.836**
		(2.49)	(7.58)
Matrix Inequality Constraints			
$\mathbf{adj(I-B)]}\boldsymbol{\omega}$	1.509	0.786	-0.252
$\mathbf{adj(I}\phi_1\mathbf{-B)A}$	0.135	0.237	-0.565
	0.001	*0.001*	*0.001*
	-0.001	-0.001	0.003
$\mathbf{BA}$	77.456	-10.543	-38.043
	0.901	*22.648*	*0.089*
	-4.599	11.168	247.392
$\mathbf{B}^2\mathbf{A}$	70.269	-7.765	-0.905
	0.810	*20.354*	*0.082*
	-3.726	11.690	206.751

Notes are as in Table 5.

8 Conclusions

In this paper we have examined some of the properties of N-dimensional extended GARCH/HEAVY models and MEM. For the parameters of these systems we have derived matrix inequality constraints that require the power transformed conditional means or variances to be almost surely non-negative at all t. Our methodology allows us to communicate such non-negativity conditions in a more user friendly way so that their implications can be seen explicitly. The conditions are not only sufficient but necessary as well. Often in practice these constraints are not taken into account. As a result many seminal papers report estimated parameters with negative values, which frequently violate the non-negativity conditions.

We have shown that the more general asymmetric setting considerably increases (actually doubles) the number of constraints and, therefore, imposes severe restrictions on the parameter space. We have also dealt with those cases where the non-negativity conditions are violated in a different way. By modelling the relationship between the conditional variables using a mixture of power and logarithmic transformations we have minimized the number of such constraints.

These findings are of interest in themselves but they also matter because they raise a number of new questions that we believe may be useful in motivating future research. Here we highlight three suggestions. Further research should try to follow our techniques and derive matrix inequality constraints ideally in multivariate systems of order higher than $(1, q)$. The second suggestion refers to a new methodological approach for obtaining explicit formulas of the second moment structure for such higher orders models. Note that He and Teräsvirta (2004) have provided only recursive solutions for the restricted extended multivariate GARCH system of order $(2, 2)$. Our less complicated procedure of adopting the ARMA representation of a GARCH model will enable us to achieve this goal. A third suggestion for future research is to relax the assumption of constant parameters. That is, to derive necessary and sufficient non-negativity conditions for N-dimensional models in a time varying setting would strengthen what we know about such systems. This is undoubtedly a difficult task, but it highlights the importance of our technique. For such 'time varying' multivariate models it is not possible to obtain univariate representations. Therefore, the approach adopted in Conrad and Karanasos (2010) is not applicable in this case. In sharp contrast, the multivariate Wold-Crámer decomposition that we have proposed in this paper, coupled with the novel methodology, proposed in Canepa et al. (2017), for dealing with 'time varying' models, can provide a solution to this problem.

References

[1] Admati, A. R. and P. C. Pfleiderer (1988). "A theory of intraday patterns: Volume and price variability." Review of Financial Studies 1(1): 3-40.

[2] Andersen, T. G., T. Bollerslev and F. X. Diebold (2007). "Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility." The Review of Economics and Statistics 89(4): 701-720.

[3] Canepa, A, M. Karanasos and A. G. Paraskevopoulos (2017). "A unified theory for the family of time varying models with ARMA representations: One solution fits all." Working Paper, Brunel University.

[4] Cipollini, F., R. Engle and G. M. Gallo (2013). "Semiparametric vector MEM." Journal of Applied Econometrics 28(7): 1067-1086.

[5] Cipollini, F. and G. M. Gallo (2010). "Automated variable selection in vector multiplicative error models." Computational Statistics & Data Analysis 54(11): 2470-2486.

[6] Conrad, C. (2010). "Non-negativity conditions for the hyperbolic GARCH model." Journal of Econometrics 157(2): 441-457.

[7] Conrad, C. and M. Karanasos (2010). "Negative volatility spillovers in the unrestricted ECCC-GARCH model." Econometric Theory 26(3): 838.

[8] de Almeida, D., L. K. Hotta and E. Ruiz (2018). "MGARCH models: Trade-off between feasibility and flexibility." International Journal of Forecasting 34(1): 45-63.

[9] Diamond, D. W. and R. E. Verrecchia (1987). "Constraints on short-selling and asset price adjustment to private information." Journal of Financial Economics 18(2): 277-311.

[10] Ding, Z., C. W. J. Granger and R. F. Engle (1993). "A long memory property of stock market returns and a new model." Journal of Empirical Finance 1(1): 83-106.

[11] Easley, D. and M. O'Hara (1987). "Price, trade size, and information in securities markets." Journal of Financial Economics 19(1): 69-90.

[12] Easley, D. and M. O'Hara (1992). "Time and the process of security price adjustment." Journal of Finance 47(2): 577-605.

[13] Engle, R. F. (2000). "The Econometrics of Ultra-high-frequency Data." Econometrica 68(1): 1-22.

[14] Engle, R. F. and G. M. Gallo (2006). "A multiple indicators model for volatility using intra-daily data." Journal of Econometrics 131(1): 3-27.

[15] Forsberg, L. and E. Ghysels (2007). "Why do absolute returns predict volatility so well?" Journal of Financial Econometrics 5(1): 31-67.

[16] Francq, C., M. D. Jiménez-Gamero and S. G. Meintanis (2017). "Tests for conditional ellipticity in multivariate GARCH models." Journal of Econometrics 196(2): 305-319.

[17] Francq, C. and G. Sucarrat (2017). "An equation-by-equation estimator of a multivariate log-GARCH-X model of financial returns." Journal of Multivariate Analysis 153: 16-32.

[18] Glosten, L. R., R. Jagannathan and D. E. Runkle (1993). "On the relation between the expected value and the volatility of the nominal excess return on stocks." The Journal of Finance 48(5): 1779-1801.

[19] Gourieroux, C. (2007). "Positivity conditions for a bivariate autoregressive volatility specification." Journal of Financial Econometrics 5(4): 624-636.

[20] Hautsch, N. (2008). "Capturing common components in high-frequency financial time series: A multivariate stochastic multiplicative error model." Journal of Economic Dynamics and Control 32(12): 3978-4015.

[21] He, C. and T. Teräsvirta (1999). "Properties of moments of a family of GARCH processes." Journal of Econometrics 92(1): 173-192.

[22] He, C. and T. Teräsvirta (2004). "An extended constant conditional correlation GARCH model and its fourth-moment structure." Econometric Theory 20(05): 904-926.

[23] Jeantheau, T. (1998). "Strong consistency of estimators for multivariate ARCH models." Econometric Theory 14(1): 70-86.

[24] Karanasos, M. and A. Kartsaklas (2009). "Dual long-memory, structural breaks and the link between turnover and the range-based volatility." Journal of Empirical Finance 16(5): 838-851.

[25] Karanasos, M, S. Yfanti and Y. Xu (2017). "Stylized facts for extended HEAVY/GARCH models and MEM: the importance of asymmetries, power transformations, hyperbolic long memory, structural breaks and spillovers." Working paper, Brunel University.

[26] Knight, J. and S. Satchell (2007). "GARCH predictions and the prediction of option prices." In. Knight, J. and S. Satchell, Eds. Forecasting Volatility in the Financial Markets. Elsevier Monographs.

[27] Li, J. and C. Wu (2006). "Daily return volatility, bid-ask spreads, and information flow: Analyzing the information content of volume." The Journal of Business 79(5): 2697-2739.

[28] Manganelli, S. (2005). "Duration, volume and volatility impact of trades." Journal of Financial Markets 8(4): 377-399.

[29] Morana, C. (2015). "Semiparametric estimation of multivariate GARCH models." Working Paper No. 317, University of Milan, Bicocca, Department of Economics.

[30] Nakatani, T. and T. Teräsvirta (2008). "Positivity constraints on the conditional variances in the family of conditional correlation GARCH models." Finance Research Letters 5(2): 88-95.

[31] Nakatani, T. and T. Teräsvirta (2009). "Testing for volatility interactions in the Constant Conditional Correlation GARCH model." The Econometrics Journal 12(1): 147-163.

[32] Nelson, D. B. (1991). "Conditional heteroskedasticity in asset returns: A new approach." Econometrica 59(2): 347-370.

[33] Nelson, D. B. and C. Q. Cao (1992). "Inequality constraints in the univariate GARCH model." Journal of Business & Economic Statistics 10(2): 229-235.

[34] Nielsen, H. B. and A. Rahbek (2014). "Unit root vector autoregression with volatility induced stationarity." Journal of Empirical Finance 29: 144-167.

[35] Noureldin, D., N. Shephard and K. Sheppard (2012). "Multivariate high-frequency-based volatility (HEAVY) models." Journal of Applied Econometrics 27(6): 907-933.

[36] Pedersen, R. S. and A. Rahbek (2014). "Multivariate variance targeting in the BEKK–GARCH model." The Econometrics Journal 17(1): 24-55.

[37] Rabemananjara, R. and J. M. Zakoian (1993). "Threshold ARCH models and asymmetries in volatility." Journal of Applied Econometrics 8(1): 31-49.

[38] Shephard, N. and K. Sheppard (2010). "Realizing the future: forecasting with high-frequency-based volatility (HEAVY) models." Journal of Applied Econometrics 25(2): 197-231

[39] Taylor, N. and Y. Xu (2017). "The logarithmic vector multiplicative error model: an application to high frequency NYSE stock data." Quantitative Finance 17(7):1021-1035.

[40] Tsai, H. and K. S. Chan (2007). "A note on non-negative ARMA processes." Journal of Time Series Analysis 28(3): 350-360.

[41] Tsai, H. and K. S. Chan (2008). "A note on inequality constraints in the GARCH model." Econometric Theory 24(3): 823-828.

[42] Wang, J. (2007). "Foreign equity trading and emerging market volatility: Evidence from Indonesia and Thailand. " Journal of Development Economics 84(2), 798-811.

A APPENDIX

In this Appendix we will present the identifiability and invertibility conditions for the SUE-AP model.

ASSUMPTIONS

Assumption A1 (Identifiability). The formulation of the N-dimensional SUE-AP model at the true values of the parameters is minimal if $\mathbf{I}-\mathbf{B}L$, $\mathbf{A}L$, and, if there are asymmetries, $(\mathbf{A}+\mathbf{\Gamma})L$ as well, satisfy the following conditions:

1. $det[\mathbf{A}L] \neq 0$ (and $det[(\mathbf{A}+\mathbf{\Gamma})L] \neq 0$, if asymmetries are present), $det[\mathbf{I}-\mathbf{B}L] \neq 0$.
2. $\mathbf{A}L$ and $\mathbf{I}-\mathbf{B}L$ are coprime. That is, any of the greatest common left divisors of $\mathbf{A}L$ and $\mathbf{I}-\mathbf{B}L$ are unimodular. In addition, if we have asymmetries, $(\mathbf{A}+\mathbf{\Gamma})L$ and $\mathbf{I}-\mathbf{B}L$ are coprime as well.
3. $\mathbf{I}-\mathbf{B}L$ is column reduced, that is $det[\mathbf{B}] \neq 0$.

Assumption A1 guarantees that the model in eq. (3) is identifiable (see also Proposition 3.4 in Jentheau, 1998 and Assumption A1 in Conrad and Karanasos, 2010).

Assumption A2 (Invertibility). The inverse roots ϕ_i, $i = 1, \ldots, N$ of $\beta(z)$ in eq. (5) lie inside the unit circle and without loss of generality are distinct and ordered as follows: $|\phi_1| > |\phi_2| > \cdots > |\phi_N|$.

B APPENDIX

In this Appendix we will present an explicit formula for the k-step-ahead optimal (in L_2 sense) linear predictor of the SUE-APM model.

OPTIMAL PREDICTORS

The k-step-ahead ($k \in \mathbf{Z}^+$, where $\mathbf{Z}^+$ is the set of positive integers) optimal (in L_2 sense) linear predictor of $\boldsymbol{\sigma}_{M,t}$, $\mathbb{E}(\boldsymbol{\sigma}_{M,t}\,|\mathcal{F}_{t-k-1})$, is readily seen to be

$$\mathbb{E}(\boldsymbol{\sigma}_{M,t}\,|\mathcal{F}_{t-k-1}) = (\mathbf{I}-\mathbf{C})^{-1}(\mathbf{I}-\mathbf{C}^k)\overline{\boldsymbol{\omega}}+\mathbf{C}^k\boldsymbol{\sigma}_{M,t-k}, \tag{B.1}$$

where $\mathbf{C} = \mathbb{E}[\mathbf{C}_t] = \mathbf{B} + (\mathbf{A} + \mathbf{\Gamma}\frac{1}{2})\mathbf{I}_{[d]}\mathbf{Z}$, and $\overline{\boldsymbol{\omega}} = \boldsymbol{\omega} + (\mathbf{A} + \mathbf{\Gamma}\frac{1}{2})\mathbf{I}_{(\mathbf{d})}\mathbf{z}$, with $\mathbf{z} = \mathbb{E}[\mathbf{z}_t]$ and $\mathbf{Z} = \mathbb{E}[\mathbf{Z}_t]$; $\mathbf{I}_{[d]}$ and $\mathbf{I}_{(d)}$ have been defined in eq. (10).
In addition, the first-order moment vector, $\boldsymbol{\sigma}_M = \mathbb{E}(\boldsymbol{\sigma}_{M,t})$, exists if and only if

$$\lambda(\mathbf{C}) < 1, \tag{B.2}$$

where $\lambda(\mathbf{C})$ refers to the modulus of the largest eigenvalue of $\mathbf{C}$. Under the condition in eq. (B.2), $\boldsymbol{\sigma}_M = \lim_{k\to\infty} \mathbb{E}(\boldsymbol{\sigma}_{M,t} \,|\mathcal{F}_{t-k-1})$, is given by

$$\boldsymbol{\sigma}_M = (\mathbf{I} - \mathbf{C})^{-1}\overline{\boldsymbol{\omega}}. \tag{B.3}$$

Similarly, the k-step-ahead optimal (in L_2 sense) linear predictor of $\boldsymbol{\varepsilon}_{M,t}$ is:

$$\mathbb{E}(\boldsymbol{\varepsilon}_{M,t} \,|\mathcal{F}_{t-k-1}) = \mathbf{I}_{[d]}\mathbf{Z}\mathbb{E}(\boldsymbol{\sigma}_t \,|\mathcal{F}_{t-k-1}) + \mathbf{I}_{(d)}\mathbf{z}. \tag{B.4}$$

When $d = N$, eq. (B.4) gives: $\mathbb{E}(\boldsymbol{\varepsilon}_t \,|\mathcal{F}_{t-k-1}) = \mathbf{Z}\mathbb{E}(\boldsymbol{\sigma}_t \,|\mathcal{F}_{t-k-1})$, where $\mathbb{E}(\boldsymbol{\sigma}_t \,|\mathcal{F}_{t-k-1})$ is obtained by eq. (B.10 by setting $d = N$.

In addition, under the condition in eq. (B.2), $\boldsymbol{\varepsilon}_M = \lim_{k\to\infty} \mathbb{E}(\boldsymbol{\varepsilon}_{M,t} \,|\mathcal{F}_{t-k-1})$, is given by

$$\boldsymbol{\varepsilon}_M = \mathbf{I}_{[d]}\mathbf{Z}(\mathbf{I} - \mathbf{C})^{-1}\overline{\boldsymbol{\omega}} + \mathbf{I}_{(d)}\mathbf{z}. \tag{B.5}$$

Notice that eq. (B.3) imposes an additional matrix inequality constraint on the parameter space, that is $(\mathbf{I} - \mathbf{C})^{-1}\overline{\boldsymbol{\omega}} > \mathbf{0}$. Finally, setting $d = N$ in eq. (B.5), we obtain $\boldsymbol{\varepsilon} = \mathbb{E}(\boldsymbol{\varepsilon}_t)$. The proofs are presented in Section C of the supplementary Appendix.

C APPENDIX

Now that we have presented the optimal predictors and the first unconditional moment of the SUE-APM model, we will examine its second moments.

SECOND MOMENTS

But first, we will introduce some additional notation, which involves various Kronecker products. Specifically, Let

$$\begin{aligned} \mathbf{Z}^{\otimes 2} &= \mathbf{Z} \otimes \mathbf{Z},\ \mathbf{Z}_*^{\otimes 2} = \mathbb{E}(\mathbf{Z}_t \otimes \mathbf{Z}_t),\ \widetilde{\mathbf{Z}} = \mathbf{Z}_*^{\otimes 2} - \mathbf{Z}^{\otimes 2}, \\ \mathbf{C}^{\otimes 2} &= \mathbf{C} \otimes \mathbf{C},\ \mathbf{C}^{\otimes I} = \mathbf{C} \otimes \mathbf{I},\ \mathbf{I}^{\otimes C} = \mathbf{I} \otimes \mathbf{C}, \\ \widetilde{\mathbf{A}} &= \mathbb{E}(\mathbf{A}_t \otimes \mathbf{A}_t), \end{aligned} \tag{C.1a}$$

and

$$\begin{aligned} \mathbf{I}_{[d]}^{\otimes 2} &= \mathbf{I}_{[d]} \otimes \mathbf{I}_{[d]},\ \mathbf{I}_{(d)}^{\otimes 2} = \mathbf{I}_{(d)} \otimes \mathbf{I}_{(d)}, \\ \mathbf{I}_{[d]}^{\otimes I_{(d)}} &= \mathbf{I}_{[d]} \otimes \mathbf{I}_{(d)},\ \mathbf{I}_{(d)}^{\otimes I_{[d]}} = \mathbf{I}_{(d)} \otimes \mathbf{I}_{[d]}. \end{aligned} \tag{C.1b}$$

Finally, denote

$$\widetilde{\mathbf{C}} = \mathbf{I}_{N^2} - \mathbf{I}^{\otimes \mathbf{C}}\mathbf{C}^{\otimes \mathbf{I}} - \widetilde{\mathbf{A}}\mathbf{I}_{[d]}^{\otimes 2}\widetilde{\mathbf{Z}}. \tag{C.1c}$$

Notice that $\widetilde{\mathbf{Z}}$ in eq. (C.1a) is a diagonal matrix (of order N^2), and its rth element, with $r = [(i-1)N + j]$, where for each $i = 1, \ldots, N$, $j = 1, \ldots, N$, is given by

$$z_{(i-1)N+j} = \mathbb{E}(|e_{it}|^{\delta_i} |e_{jt}|^{\delta_j}) - \mathbb{E}(|e_{it}|^{\delta_i})\mathbb{E}(|e_{jt}|^{\delta_j}).$$

Therefore, $\widetilde{\mathbf{Z}}\mathbf{j}_{N^2}$ is a vector of order N^2 with $[(i-1)N + j]$th element $z_{(i-1)N+j}$.

Observe also that the four matrices in eq. (C.1b), are diagonal of order N^2, with only their $[(l-1)N+i]$th diagonal elements non-zero, where (for $d \neq 0, N$):

For $\mathbf{I}_{[d]}^{\otimes 2}$: $l, i = 1, \ldots, d, \quad d \geq 1,$
For $\mathbf{I}_{(d)}^{\otimes 2}$: $l, i = d+1, \ldots, N, \quad d \leq N-1,$
For $\mathbf{I}_{[d]}^{\otimes I_{(d)}}$: $l = 1, \ldots, d,\ d \geq 1\ ;\ i = d+1, \ldots, N,\ d \leq N-1,$
For $\mathbf{I}_{(d)}^{\otimes I_{[d]}}$: $l = d+1, \ldots, N,\ d \leq N-1;\ i = 1, \ldots, d, \quad d \geq 1.$

Next, let $\mathbf{\Gamma}_M(l) = [\gamma_{M,ij}(l)]$, $l \in \mathbb{Z}$, denote the multidimensional covariance function of $\{\boldsymbol{\sigma}_{M,t}\}$, that is

$$\mathbf{\Gamma}_M(l) = \mathbb{E}[(\boldsymbol{\sigma}_{M,t-l} - \boldsymbol{\sigma}_M)(\boldsymbol{\sigma}_{M,t} - \boldsymbol{\sigma}_M)'],$$

or

$$\mathbf{\Gamma}_M(l) = \mathbf{\Sigma}_M(l) - \boldsymbol{\sigma}_M\ \boldsymbol{\sigma}'_M,$$

where $\mathbf{\Sigma}_M(l) = \mathbb{E}(\boldsymbol{\sigma}_{M,t-l}\boldsymbol{\sigma}'_{M,t})$. In addition, let the vec forms of $\mathbf{\Sigma}_M(l)$ and $\mathbf{\Gamma}_M(l)$ denoted by $\mathbf{s}_M(l)$ and $\boldsymbol{\gamma}_M(l)$, respectively. Explicit solutions for the $\mathbf{\Gamma}_M(l)$ and conditions for its existence will be presented below.

Further, let

$$\mathbf{D}_M = diag[\sqrt{\gamma_{M,11}(0)}, \ldots, \sqrt{\gamma_{M,NN}(0)}],$$

where $\gamma_{M,ii}(0)$ is the ith diagonal element of $\mathbf{\Gamma}_M(0)$. To further fix notation, write the lth-order, for $l \geq 1$, autocorrelation matrix of $\{\boldsymbol{\sigma}_{M,t}\}$ as

$$\mathbf{R}_M(l) = \mathbf{D}_M^{-1}\mathbf{\Gamma}_M(l)\mathbf{D}_M^{-1}.$$

Assumption C1. $\lambda_{\max}(\widetilde{\mathbf{C}}) < 1$.

Under Assumption C1 ($\widetilde{\mathbf{C}}$ has been defined in eq. (C.1c)) the vec form of $\mathbf{\Gamma}_M(0)$ is given by

$$\boldsymbol{\gamma}_M(0) = \widetilde{\mathbf{C}}^{-1}\widetilde{\mathbf{A}}(\mathbf{I}_{[d]}^{\otimes 2}\widetilde{\mathbf{Z}}\boldsymbol{\sigma}^{\otimes 2} + \mathbf{I}_{(d)}^{\otimes 2}\widetilde{\mathbf{Z}}\mathbf{j}_{N^2} + \mathbf{I}_{[d]}^{\otimes \mathbf{I}_{(d)}}\widetilde{\mathbf{Z}}\boldsymbol{\sigma}^{\otimes j} + \mathbf{I}_{(d)}^{\otimes \mathbf{I}_{[d]}}\widetilde{\mathbf{Z}}\mathbf{j}^{\otimes \sigma}). \tag{C.2}$$

Further, the vec form of the covariance function, for lag $l \geq 1$, $\boldsymbol{\gamma}_M(l)$, is given by

$$\boldsymbol{\gamma}_M(l) = \mathbf{I}^{\otimes \mathbf{C}}\boldsymbol{\gamma}_M(0).$$

Notice that eq. (C.2) imposes an additional matrix inequality constraint on the parameter space, that is $\mathbf{D}_M > \mathbf{0}$.

The proofs are presented in Section E of the supplementary Appendix.