Constrained QML Estimation for Multivariate Asymmetric MEM with Spillovers: The Importance of Matrix Inequalities

M. Karanasos[a], Y. Xu[b], S. Yfanti[c]
[a] Brunel University London, UK;[b] Cardiff University, UK;
[c] Loughborough University, UK;

April 7, 2022

Abstract

In this paper we review and generalize results on the derivation of tractable non-negativity (necessary and sufficient) conditions for N-dimensional asymmetric MEM and GARCH/HEAVY models with spillovers. We show that the non-negativity constraints are translated into simple matrix inequalities, which are easily handled. In practice these conditions may not be fulfilled. To deal with these cases we propose a constrained QML estimation. 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: Multivariate Statistics; Constrained QML Estimation; Matrix Inequalities; Multivariate Modelling; Second Moment Structure

JEL Classification: C32, C53, C58, G15.

Declaration of interest: none.

Author details:

Corresponding author: Professor Menelaos Karanasos, Economics and Finance, Brunel University London, Kingston Lane, Uxbridge, UB8 3PH, UK; e-mail: menelaos.karanasos@brunel.ac.uk, tel: +44 (0)1895 265284.

Dr. Yongdeng Xu, Cardiff Business School, Cardiff University, Aberconway Building, Colum Drive, Cardiff, CF10 3EU, UK; e-ail: xuy16@cardiff.ac.uk.

Dr. Stavroula Yfanti, School of Business and Economics, Loughborough University, Loughborough, Leicestershire, LE11 3TU, UK; e-mail: stavyfan@gmail.com, tel: +44 (0) 1509227091.

1 Introduction

Multivariate MEM (multiplicative error models) and GARCH/HEAVY[1] processes have found considerable empirical success (see, for example, Christodoulakis and Satchell, 2002, Christodoulakis, 2007, Nakatani and Teräsvirta, 2009, Shephard and Sheppard, 2010, Conrad and Karanasos, 2010, Noureldin et al., 2012, Cipollini et al., 2013, Pedersen and Rahbek, 2014, Morana, 2015, Taylor and Hu, 2017, **MORE RECENT REFERENCES ON MULTIV MEM**).[2]

As was indicated at the outset of the research by Conrad and Karanasos (2010) the standard restricted extended multivariate GARCH model, introduced by Jeantheau (1998), 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 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 MEM and GARCH/HEAVY processes. 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 a number of papers report estimated coefficients whose values violate the non-negativity conditions (see Section D of the supplementary Appendix).

Second, we also derive new tractable constraints for the asymmetric versions of these N-dimensional systems. These allow new matrix inequalities to be constructed for the asymmetric 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.

Apart from conditions on multivariate asymmetric MEM and GARCH models, which ensure the non-negativity of the conditional variables, we also derive (our third contribution) theoretical results on the optimal forecasts and on the second order moments of such models, and, therefore, we complement Karanasos (2000) and 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

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

[2] For other recent developments in multivariate GARCH models see, for example, Nielsen and Rahbek (2014), and de Almeida et al. (2018).

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 even effortlessly enforced in estimation. Thus, a final contribution is the consistent estimation of the multivariate MEM by constrained quasi maximum likelihood (QML) estimation instead of the efficient generalized method of moments estimation used in Cipollini et al. (2013). A Monte Carlo simulation confirms the superiority of the constrained estimation over the unconstrained one (see Section 5). 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 6), 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.

It is important to highlight the fact that our methodology is very general, and it can be applied not only in multivariate MEM and GARCH/HEAVY processes, but also to vector AR (VAR) models where the variables in the system should take non-negative values (see, for example, **REFERENCES**). That is, our matrix inequality constraints should be enforced to these models as well.

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 specification. Section 3 reviews the model and its “one-sided” representation, and tractable expressions for the non-negativity constraints are presented together with some numerical examples in Section 4. The next Section presents the Monte Carlo simulation. Section 6 contains the empirical examples, and the conclusions can be found in Section 7. The Appendix briefly discusses the optimal forecasts and the second moment structure of our model. An online Appendix (available online) contains the proofs.

2 Preliminaries

2.1 Notation

Throughout the paper we will adhere to the following notation.

Notation 1 *Throughout the paper we adhere to the following conventions: the set of integers (resp. positive and non-negative integers) is denoted by* $\mathbb{Z}$ *(resp.* $\mathbb{Z}_{>0}$ *and* $\mathbb{Z}_{\geq 0}$*). Similarly, the set of real numbers (resp. positive and non-negative) is denoted by* $\mathbb{R}$ *(resp.* $\mathbb{R}_{>0}$ *and* $\mathbb{R}_{\geq 0}$*).*
We will use upper(lower) case boldface symbols to refer to square matrices(vectors). That is, $\mathbf{y} = [y_i]_{i=1,\ldots,N}$ *is an* $N \times 1$ *column vector,* $N \in \mathbb{Z}_{\geq 1}$*,* $\mathbf{Y} = [y_{ij}]_{i,j=1,\ldots,N}$ *is a square matrix of order* N *(hereafter we will drop the subscript for notational simplicity).*
Let $(\Omega, \mathcal{F}, P)$ *be a probability space and* $L_2(\Omega, \mathcal{F}, P)$ *(in short* L_2*) be the Hilbert space of random variables with finite first and second moments defined on* $(\Omega, \mathcal{F}, P)$*.*

Notation 2 *The elementwise expectation operator is denoted by* $\mathbb{E}$*, i.e.,* $\mathbb{E}(\mathbf{y}) = [\mathbb{E}(y_i)]$*. Similarly,* $\mathbb{E}(\mathbf{y}_t\,|\mathcal{F}_{t-1}\,)$ *denotes the elementwise, conditional on time* $t-1$*, expectation given* $\mathcal{F}_{t-1}$*, where* $\mathcal{F}_{t-1}$ *is the smallest closed subspace of* L_2 *based on the sequence of past observations* $\{\mathbf{y}_{t-1}, \mathbf{y}_{t-2}, \ldots\}$*, which also contains all constant functions.*
The identity matrix of order N *is denoted by* $\mathbf{I}_N$*,* $\mathbf{0}_N$ *is the null matrix, and* $\mathbf{j}_N$ *is a vector of ones. Hereafter, for notational simplicity we will drop the subscript when the order is* N*.*
Using standard notation, $\mathbf{Y}'$ *and* $\mathbf{Y}^{-1}$ *are the transpose and the inverse of the square matrix* $\mathbf{Y}$*. In addition,* $\mathbf{Y} = diag[\mathbf{y}]$ *is a diagonal matrix with* y_i *occupying the i-th diagonal entry. Finally, the inequality* $\mathbf{Y} \geq \mathbf{0}$ *means that all elements of* $\mathbf{Y}$ *are non-negative real numbers.*

Notation 3 *Superscripts within parentheses or brackets (e.g.,* $(.)^{(m)}$*) designate the index position of the corresponding term (e.g., m-th term) of a sequence, so as to distinguish position indices from power exponents.*
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 j-th row and i-th column.*

In other words, $Y_{ij}^{\{a\}}$ *is the cofactor of the ji-th element of* $\mathbf{Y}$*.*
Let also $Y_{ij}(L)$ *be a polynomial (i.e., of order* N*) where* L *is the lag operator. Then* $\mathbf{Y}(L) = [Y_{ij}(L)]$ *indicates a matrix polynomial in the lag operator.*

Notation 4 $\mathbf{Y}^{\wedge k} = [y_{ij}^k]$ *is the elementwise exponentiation, whereas* $\mathbf{y}^{\wedge \mathbf{x}} = [y_i^{x_i}]$*, that is the element occupying the i-th entry of vector* $\mathbf{y}$ *is raised to the power of the element occupying the i-th entry of vector* $\mathbf{x}$*.* $\mathbf{Y}^k = \prod_{i=1}^k \mathbf{Y}$ *means that the matrix* $\mathbf{Y}$ *is raised to the power of* k*.*
We will refer to the elementwise absolute value of $\mathbf{Y}$ *as* $abs[\mathbf{Y}] = [|y_{ij}|]$*.* $\max[\mathbf{Y}]$ *indicates the largest element of the matrix* $\mathbf{Y}$*.* $\log[\mathbf{Y}]$ *means that we take the log of each element of* $\mathbf{Y}$*, that is* $\log[\mathbf{Y}] = [\log(y_{ij})]$*.*

Notation 5 *Let* $\mathbf{Y}^{\otimes 2} = \mathbf{Y} \otimes \mathbf{Y}$*, where* $\otimes$ *is the Kronecker product of two matrices. In addition,* $vec(\mathbf{Y})$ *is a vector in which the columns of the matrix* $\mathbf{Y}$ *are stacked one underneath the other.*

2.2 The Model

In this Section we introduce the asymmetric (with spillovers) specification. Consider the N-dimensional vector process, $\boldsymbol{\varepsilon}_t = [\varepsilon_{it}]$. We assume that the vector $\boldsymbol{\varepsilon}_t$ is characterized by the relation

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

where $\mathbf{Z}_t = diag[\mathbf{e}_t]$, $\mathbf{e}_t = [e_{it}]$, and $\boldsymbol{\sigma}_t = [\sigma_{it}]$ is $\mathcal{F}_{t-1}$ measurable. That is, $\boldsymbol{\varepsilon}_t = [e_{it}\sigma_{it}]$. The stochastic vector $\mathbf{e}_t$ is independent and identically distributed $(i.i.d)$.

In the N-dimensional MEM the vector $\mathbf{e}_t$ is positive: $\mathbf{e}_t > \mathbf{0}_{N\times 1}$. Its conditional expected value is the unit vector: $\mathbb{E}(\mathbf{e}_t \,|\mathcal{F}_{t-1}\,) = \mathbf{j}$, and, thus $\mathbb{E}(\boldsymbol{\varepsilon}_t \,|\mathcal{F}_{t-1}\,) = \boldsymbol{\sigma}_t$. Moreover, it has positive definite conditional covariance matrix $\mathbf{Q} = [q_{ij}]$, with: $\mathbf{q} = [q_{ii}]$, that is $\mathbf{q}$ is the vector with the diagonal elements of $\mathbf{Q}$, and we also denotes as $\widetilde{\mathbf{Q}} = diag[\mathbf{q}]$, that is $\widetilde{\mathbf{Q}}$ is equal to the $\mathbf{Q}$ matrix with its off-diagonal elements equal to zero. Therefore, the conditional correlation matrix of $\mathbf{e}_t$, denoted with $\mathbf{R} = [\rho_{ij}]$, is given by $\mathbf{R} = \widetilde{\mathbf{Q}}^{-1/2}\mathbf{Q}\widetilde{\mathbf{Q}}^{-1/2}$, that is $\rho_{ij} = q_{ij}/\sqrt{q_{ii}q_{jj}}$.

Next let us denote: as $\boldsymbol{\Sigma}_t^* = diag[\boldsymbol{\sigma}_t]$, by $\boldsymbol{\Sigma}_t = [\sigma_{ij,t}]$ the conditional covariance matrix of $\boldsymbol{\varepsilon}_t$, and with $\widetilde{\boldsymbol{\Sigma}}_t$ the $\boldsymbol{\Sigma}_t$ matrix with its off-diagonal elements equal to zero. It is straightforward to show that $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t^*\mathbf{Q}\boldsymbol{\Sigma}_t^*$, that is $\sigma_{ij,t} = \sigma_{it}\sigma_{jt}q_{ij}$, and that the conditional correlation of $\boldsymbol{\varepsilon}_t$ is also $\mathbf{R}$.[3]

A major problem in specifying a valid multivariate MEM lies in choosing appropriate parametric specifications such that $\boldsymbol{\Sigma}_t$ is positive definite almost surely for all t. Positive definiteness of $\boldsymbol{\Sigma}_t$ follows if the covariance matrix $\mathbf{Q}$ being positive definite, and $\boldsymbol{\sigma}_t > \mathbf{0}_{N\times 1}$ (if the conditional mean of ε_{it}, that is σ_{it}, is positive for all i and t).[4]

The N-dimensional multivariate unrestricted full asymmetric (MUFA) MEM of order $(1, q)$ consists of the following equations:

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

where $q \in \mathbb{Z}_{\geq 1}$, s_{jt} is a dummy variable that: i) in the case where j represents the index of stock returns, it takes the value 1 if $e_{jt} < 0$, 0 otherwise, that is $s_{jt} = 0.5[1 - sign(e_{jt})1]$, for all t, and ii) in the case where j represents the index of the squared signed rooted (SSR) realized measure of volatility, it takes the value 1 if a signed variable (i.e., the underlying stock returns) $x_{jt} < 0$, 0 otherwise, for all t.[5]

[3] In the N-dimensional (constant conditional correlation) GARCH model $\mathbf{e}_t$ has zero conditional mean and unit conditional variance, that is $\mathbb{E}(\boldsymbol{\varepsilon}_t^2 \,|\mathcal{F}_{t-1}\,) = \boldsymbol{\sigma}_t^{\wedge 2}$. In addition, it has positive definite (time invariant) conditional correlation matrix $\mathbf{R} = [\rho_{ij}]$ with $\rho_{ii} = 1$. In other words, the conditional covariance matrix of $\boldsymbol{\varepsilon}_t$, that is $\boldsymbol{\Sigma}_t = \mathbb{E}(\boldsymbol{\varepsilon}_t\boldsymbol{\varepsilon}_t' \,|\mathcal{F}_{t-1}\,)$, is given by $\boldsymbol{\Sigma}_t = \boldsymbol{\Sigma}_t^*\mathbf{R}\boldsymbol{\Sigma}_t^*$, where $\boldsymbol{\Sigma}_t^* = diag[\boldsymbol{\sigma}_t]$.

[4] Similarly, for the GARCH model we need in addition to the correlation matrix $\mathbf{R}$ being positive definite, the conditional variances of ε_{it}, that is σ_{it}^2, to be positive (for all i and t) as well.

[5] This type of asymmetry was introduced by Glosten et. al. (1993). We also consider a second type of asymmetry, which was introduced by Ding et. al. (1993), and we refer to it as MUFA model 2 (results are available upon request).

Remark 1 *The MUFA process can be either a multivariate MEM or a GARCH/HEAVY model. For example, in the bivariate context the two variables, ε_{it}, $i = 1, 2$, can be the squared returns and the realized measure, and $\sigma_{it} = E(\varepsilon_{it} | \mathcal{F}_{t-1})$ their conditional means. In the latter model the two GARCH variables (ε_{it}) can be the stock returns and the SSR realized measure (i.e., realized variance), and $\sigma^2_{it} = E(\varepsilon^2_{it} | \mathcal{F}_{t-1})$ their conditional variances. The HEAVY formulation parallels the GARCH one. Therefore the three terms, MEM, GARCH, HEAVY, can be used* interchangeably *in our analysis.*

The MUFA MEM$(1, q)$ can be expressed/interpreted as an N-dimensional system with shock (unconditional) and conditional spillovers[6]:

$$(\mathbf{I} - \mathbf{B}L)\boldsymbol{\sigma}_t = \boldsymbol{\omega} + \sum_{l=1}^{q} L^l \mathbf{A}_t^{(l)} \boldsymbol{\varepsilon}_t, \tag{2}$$

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; $\mathbf{A}_t^{(l)} = \mathbf{A}^{(l)} + \boldsymbol{\Gamma}_t^{(l)}$, where $\mathbf{A}^{(l)} = [\alpha_{ij}^{(l)}]$ and $\boldsymbol{\Gamma}_t^{(l)} = [\boldsymbol{\gamma}_{ij}^{(l)} s_{jt}]$, $l = 1, \ldots, q$, are full matrices as well. Note that $\boldsymbol{\Gamma}_t^{(l)}$ can be written as $\boldsymbol{\Gamma}_t^{(l)} = \boldsymbol{\Gamma}^{(l)} diag[\mathbf{s}_t]$ where $\boldsymbol{\Gamma}^{(l)} = [\gamma_{ij}^{(l)}]$ and $\mathbf{s}_t = [s_{jt}]$. The cross diagonal elements of $\mathbf{A}^{(l)}$ capture the shock (or unconditional) spillovers, whereas those of $\boldsymbol{\Gamma}^{(l)}$ capture the asymmetric shock spillovers.

Remark 2 *The above N-dimensional system is termed full since all the square matrices are full, and unrestricted because as we will see below some of the elements of the* $\mathbf{B}$ *(including some of the off-diagonal ones but not all),* $\mathbf{A}^{(l)}$ *and* $\boldsymbol{\Gamma}^{(l)}$ *matrices are allowed to take not only positive but negative values as well. That is, we consider a formulation that allows for feedback effects between the conditional means/variances, which can be of either sign, positive or negative (see Theorem 1 below).*[7] *Note that, if there are no asymmetries, that is $\boldsymbol{\Gamma}^{(l)} = 0$ for all l, then the model reduces to the symmetric specification.*

Next, we will present the identifiability condition for the MUFA system.

Assumption A1 (Identifiability). The formulation of the N-dimensional MUFA specification of order $(1, q)$ at the true values of the parameters is minimal if $\mathbf{I} - \mathbf{B}L$, $\sum_{l=1}^{q} \mathbf{A}^{(l)} L^l$ (positive errors), and $\sum_{l=1}^{q} (\mathbf{A}^{(l)} + \boldsymbol{\Gamma}^{(l)}) L^l$ (negative errors), satisfy the following conditions:

1. $\mathbf{I} - \mathbf{B}L$ is column reduced, that is $det[\mathbf{B}] \neq 0$.
2. $det[\mathbf{I} - \mathbf{B}L] \neq 0$, $det[\sum_{l=1}^{q} \mathbf{A}^{(l)} L^{(l)}] \neq 0$, $det[\sum_{l=1}^{q} (\mathbf{A}^{(l)} + \boldsymbol{\Gamma}^{(l)}) L^l] \neq 0$.
3. $\sum_{l=1}^{q} \mathbf{A}^{(l)} L^l$ (the case of positive errors) and $\mathbf{I} - \mathbf{B}L$ are coprime. That is, any of the greatest common left divisors of $\sum_{l=1}^{q} \mathbf{A}^{(l)} L^l$ and $\mathbf{I} - \mathbf{B}L$ are unimodular. In addition, $\sum_{l=1}^{q} (\mathbf{A}^{(l)} + \boldsymbol{\Gamma}^{(l)}) L^l$ (negative errors) and $\mathbf{I} - \mathbf{B}L$ are coprime as well.

Assumption A1 guarantees that the model in eq. (2) is identifiable (see also, for the symmetric case, Proposition 3.4 in Jentheau, 1998 and Assumption A1 in Conrad and Karanasos, 2010).

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

3 "One-sided" Representation

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 in their paper) MEM (GARCH in their paper) 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

[6] For the GARCH model we replace $\boldsymbol{\sigma}_t$ and $\boldsymbol{\varepsilon}_t$ by $\boldsymbol{\sigma}_t^{\wedge 2}$ and $\boldsymbol{\varepsilon}_t^{\wedge 2}$, respectively.

[7] In the restricted full (symmetric) formulation of order $(1, 1)$ (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.

sufficient) conditions. The first step, is the "univariate" representations, which each conditional mean (or variance in the case of the GARCH model) admits. The second step is the infinite expansions, in terms of convolutions of infinite-order kernels and corresponding 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 online 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 MUFA system as matrix inequalities (see Theorem 1 below).

3.1 Definitions

In this Section we will present a useful Proposition. In particular, we obtain the MUFA "one-sided" infinite-order expansion of each conditional mean/variance in terms of convolutions of MEM/GARCH kernels and corresponding errors.

First, we introduce the following definitions.

Definition 1 *i) Set $\boldsymbol{\mu} = [\mu_i]$ as*

$$\boldsymbol{\mu} = adj[\mathbf{I} - \mathbf{B}]\boldsymbol{\omega}. \tag{3}$$

ii) 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

$$\beta(L) = \det[\mathbf{I} - \mathbf{B}L], \tag{4}$$

where ϕ_i are the roots of $\beta(z^{-1})$.

Under Assumption A1: $\beta_N \neq 0$, and, thus $\beta(L)$ is a scalar polynomial of order N. In what follows, without loss of generality, we will assume that ϕ_i are distinct and they satisfy the inequalities: $|\phi_1| > |\phi_2| > \cdots > |\phi_N|$ (see also Nelson and Cao, 1992).

Definition 2 *We also define the square matrix polynomial $\boldsymbol{\alpha}_t(L) = [\alpha_{ij,t}(L)]$ with $\alpha_{ij,t}(L) = \sum_{n=1}^{Nq} L^n \alpha_{ij,t}^{(n)}$, for $i, j = 1, \ldots, N$, as:*

$$\boldsymbol{\alpha}_t(L) = adj[\mathbf{I} - \mathbf{B}L] \sum_{l=1}^{q} (\mathbf{A}^{(l)} + L^l \boldsymbol{\Gamma}^{(l)} diag[\mathbf{s}_t]). \tag{5}$$

Notice that $\boldsymbol{\alpha}_t(L)$ is time varying only because of the presence of asymmetries. Similarly to $\beta(L)$, since under Assumption A1: $\alpha_{ij,t}^{(Nq)} \neq 0$ for all i, j and t, the scalar polynomials $\alpha_{ij,t}(L)$ are of order Nq.

Example 1 *For illustrative purposes consider the bivariate UFA MEM$(1, 1)$. That is, $N = 2$ and $q = 1$. For this specification $\boldsymbol{\alpha}_t(L)$ in eq. (5) in Definition 2 is a 2×2 matrix polynomial, given by*

$$\boldsymbol{\alpha}_t(L) = \begin{bmatrix} \alpha_{11,t}(L) & \alpha_{12,t}(L) \\ \alpha_{21,t}(L) & \alpha_{22,t}(L) \end{bmatrix} = \begin{bmatrix} L\alpha_{11,t}^{(1)} + L^2\alpha_{11,t}^{(2)} & L\alpha_{12,t}^{(1)} + L^2\alpha_{12,t}^{(2)} \\ L\alpha_{21,t}^{(1)} + L^2\alpha_{21,t}^{(2)} & L\alpha_{22,t}^{(1)} + L^2\alpha_{22,t}^{(2)} \end{bmatrix},$$

with

$$\begin{aligned}
\alpha_{11,t}^{(1)} &= \alpha_{11} + \gamma_{11}s_{1t},\ \alpha_{11,t}^{(2)} = (\alpha_{21}\beta_{12} - \alpha_{11}\beta_{22}) + (\gamma_{21}\beta_{12} - \gamma_{11}\beta_{22})s_{1t}, \\
\alpha_{22,t}^{(1)} &= \alpha_{22} + \gamma_{22}s_{2t},\ \alpha_{22,t}^{(2)} = (\alpha_{12}\beta_{21} - \alpha_{22}\beta_{11}) + (\gamma_{12}\beta_{21} - \gamma_{22}\beta_{11})s_{2t}, \\
\alpha_{12,t}^{(1)} &= \alpha_{12} + \gamma_{12}s_{2t},\ \alpha_{12,t}^{(2)} = (\alpha_{22}\beta_{12} - \alpha_{12}\beta_{22}) + (\gamma_{22}\beta_{12} - \gamma_{12}\beta_{22})s_{2t}, \\
\alpha_{21,t}^{(1)} &= \alpha_{21} + \gamma_{21}s_{1t},\ \alpha_{21,t}^{(2)} = (\alpha_{11}\beta_{21} - \alpha_{21}\beta_{11}) + (\gamma_{11}\beta_{21} - \gamma_{21}\beta_{11})s_{1t}.
\end{aligned}$$

Next, we will present the invertibility condition for the N-dimensional MUFA system.

Assumption A2 (Invertibility). The inverse roots ϕ_i, $i = 1, \ldots, N$ of $\beta(z)$ in eq. (4) lie inside the unit circle.

Assumption A2 guarantees that the model in eq. (2) is invertible (see also Proposition 3.4 in Jentheau, 1998 and Assumption A1 in Conrad and Karanasos, 2010).

Definition 3 *Let the invertibility condition hold, and set* $\mathbf{\Psi}_t(L) = [\Psi_{ij,t}(L)]$, *where* $\Psi_{ij,t}(L) = \sum_{k=1}^{\infty} L^k \psi_{ij,t}^{(k)}$, *as*

$$\mathbf{\Psi}_t(L) = \boldsymbol{\alpha}_t(L)/\beta(L), \tag{6}$$

and, thus

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

(recall that $\beta(L)$ *and* $\boldsymbol{\alpha}_t(L)$ *have been defined in eqs. (4) and (5), respectively). Equivalently* $\mathbf{\Psi}_t(L)$ *can be written as*

$$\mathbf{\Psi}_t(L) = \sum_{k=1}^{\infty} L^k \mathbf{\Psi}_t^{(k)}, \tag{8}$$

with $\mathbf{\Psi}_t^{(k)} = [\psi_{ij,t}^{(k)}]$.

3.2 "One-sided" Decomposition

The following Proposition gives the "one-sided" representation of the MUFA specification of order $(1, q)$. The proof is trivial (the result is obtained from eq. (2) by inverting the matrix polynomial $\mathbf{I} - \mathbf{B}L$ and using the notation in eqs. (3), (4), (5) and (7)).

Proposition 1 *Let Assumptions (A1) and (A2) be satisfied. Then the MUFA MEM*$(1, q)$ *in eq. (2) admits the multivariate "one-sided" representation:*

$$\boldsymbol{\sigma}_t = \frac{\boldsymbol{\mu}}{\beta(1)} + \frac{\boldsymbol{\alpha}_t(L)}{\beta(L)} \boldsymbol{\varepsilon}_t, \tag{9}$$

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

$$\sigma_{it} = \frac{\mu_i}{\beta(1)} + \sum\nolimits_{j=1}^{N} \Psi_{ij,t}(L) \varepsilon_{jt}. \tag{10}$$

(recall that $\Psi_{ij,t}(L)$ *has been defined in eq. (7)).*

The proof is straightforward and therefore omitted.

Remark 3 *Each* $\Psi_{ij,t}(L)$ *can be thought of as an infinite-order kernel of a "univariate" unrestricted asymmetric* (N, Nq) *specification. Clearly, for the* N*-dimensional process in eq. (2) to be well-defined and the* N *conditional variables to be positive almost surely for all* t*, all the constants* μ_i *(see eq. (3)) must be positive and all the* $\psi_{ij,t}^{(k)}$ *coefficients in the "univariate one-sided" representations, that is eq. (10), must be non-negative:* $\psi_{ij,t}^{(k)} \geq 0$, $i, j = 1, \ldots, N$ *(or equivalently,* $\mathbf{\Psi}_t^{(k)} \geq 0$*; see eq. (8)) for all* t *and* $k = 1, 2, \ldots$.

In other words, the non-negativity of the 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 MEM/GARCH kernels of the MUFA system (see also Conrad and Karanasos, 2010, who paid special attention only to the bivariate case of order $(1, 1)$).

Alternative "One-Sided" Representation

Next we make a general observation that will be applied tactically later on. That is, we shall make use of the following Corollary. The multivariate "one-sided" decomposition has been presented above. Here we present an alternative form for such an infinite-order expansion.

Corollary 1 *Let Assumptions (A1) and (A2) be satisfied. Then eq. (9) can be rewritten in an alternative form as:*

$$\boldsymbol{\sigma}_t = \frac{\boldsymbol{\mu}}{\beta(1)} + \sum_{k=1}^{\infty} L^k \boldsymbol{\Psi}_t^{(k)} \boldsymbol{\varepsilon}_t, \tag{11}$$

where

$$\boldsymbol{\Psi}_t^{(k)} = \sum_{s=1}^{\min(q,k)} \mathbf{B}^{k-s} \mathbf{A}_t^{(s)}$$

(recall that $\mathbf{A}_t^{(s)}$ has been defined in eq. (2)).

The above Corollary follows directly from Proposition 1 since:
i) $(\mathbf{I} - \mathbf{B}L)^{-1}\boldsymbol{\omega} = \frac{adj[\mathbf{I}-\mathbf{B}]\boldsymbol{\omega}}{\det[\mathbf{I}-\mathbf{B}]} \overset{(\text{Def. }1)}{=} \frac{\boldsymbol{\mu}}{\beta(1)}$, and
ii) $\boldsymbol{\Psi}_t(L) \overset{(\text{eq. }(6))}{=} \frac{\boldsymbol{\alpha}_t(L)}{\beta(L)} \overset{(\text{eq. }(5))}{=} \frac{adj[\mathbf{I}-\mathbf{B}L]}{\beta(L)} \sum_{l=1}^{q} L^l \mathbf{A}_{t-l}^{(l)} = [\mathbf{I} - \mathbf{B}L]^{-1} \sum_{l=1}^{q} L^l \mathbf{A}_{t-l}^{(l)}$,
and, hence, $\boldsymbol{\Psi}_t^{(k)}$ in eq. (8) is given by: $\boldsymbol{\Psi}_t^{(k)} = \sum_{s=1}^{\min(q,k)} \mathbf{B}^{k-s}\mathbf{A}_t^{(s)}$.

4 Matrix Inequality Constraints

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 MUFA systems easily solvable and downright tractable.

As mentioned in the previous Section checking a *finite* number of inequality constraints, that is $\boldsymbol{\Psi}_t^{(k)} \geq 0$ for a large enough k, ensures the non-negativity of all MEM/GARCH kernels of the MUFA specification of order $(1, q)$. It will suffice to show that the constraints are satisfied in the two extreme cases:

i) all the N *i.i.d* errors are positive for every t: $\mathbf{e}_t > \mathbf{0}_{Nx1}$ for all t, and thus $\mathbf{A}_t^{(l)} = \mathbf{A}^{(l)}$ in eq. (2) for all l, since $\mathbf{s}_t = 0$ for all t,

ii) for every t $\mathbf{A}_t^{(l)} = \mathbf{A}^{(l)} + \boldsymbol{\Gamma}^{(l)}$, that is, $\mathbf{e}_t < \mathbf{0}_{Nx1}$ for all t and, therefore, $\mathbf{s}_t = \mathbf{j}$. All other cases where some of the errors are positive and the rest are negative are covered by these two extreme cases.

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 (recall that the superscript with parentheses denotes an index).

Notation 6 *i) Let $\mathbf{G}^+ = [g_{ij}^+]$, used in Condition (C2a) in Theorem 1(B) below, be given by $\mathbf{G}^+ = \sum_{l=1}^{q} \mathbf{A}^{(l)} \phi_1^{q-l}$,*
ii) We also let $\mathbf{G}^- = [g_{ij}^-]$ be given by $\mathbf{G}^- = \sum_{l=1}^{q} \left(\mathbf{A}^{(l)} + \boldsymbol{\Gamma}^{(l)}\right) \phi_1^{q-l}$.
For $q = 1$ (the model of order $(1,1)$) we have: $g_{ij}^+ = \alpha_{ij}^{(1)}$ and $g_{ij}^- = \alpha_{ij}^{(1)} + \gamma_{ij}^{(1)}$.

Notation 7 *i) Similarly to Notation 6, let $\boldsymbol{\Psi}^{+(k)} = [\psi_{ij}^{+(k)}]$ be given by $\boldsymbol{\Psi}^{+(k)} = \sum_{s=1}^{\min(q,k)} \mathbf{B}^{k-s}\mathbf{A}^{(s)}$ (see Condition (C3a) in Theorem 1(B) below). That is $\boldsymbol{\Psi}_t^{(k)}$ (given in Corollary 1) is equal to $\boldsymbol{\Psi}^{+(k)}$ when $\mathbf{e}_t > \mathbf{0}_{Nx1}$, or equivalently, when we have the symmetric case: $\mathbf{s}_t$ is a zero vector,*
ii) We will also use the notation $\boldsymbol{\Psi}^{-(k)} = [\psi_{ij}^{-(k)}]$ where $\boldsymbol{\Psi}^{-(k)} = \sum_{s=1}^{\min(q,k)} \mathbf{B}^{k-s} \left(\mathbf{A}^{(s)} + \boldsymbol{\Gamma}^{(s)}\right)$. In other words $\boldsymbol{\Psi}_t^{(k)} = \boldsymbol{\Psi}^{-(k)}$ when $\mathbf{e}_t < \mathbf{0}_{Nx1}$, or equivalently, $\mathbf{s}_t = \mathbf{j}$ (only negative errors).
For $k = 1$, we have: $\psi_{ij}^{+(1)} = \alpha_{ij}^{(1)}$ and $\psi_{ij}^{-(1)} = \alpha_{ij}^{(1)} + \gamma_{ij}^{(1)}$.

Notation 8 *i) Let* $\mathbf{Y}^{(n)} = [y_{ij}^{(n)}]$, $n = 1, \dots, N$, *and* $\max[\mathbf{Y}^{(n)}] = \max(y_{ij}^{(n)})$ *for* $i, j = 1, \dots, N$. *In other words,* $\max[\mathbf{Y}^{(n)}]$ *is the largest element of all the* N^2 *elements of* $\mathbf{Y}^{(n)}$,
ii) $\mathbf{Y}_{\max}^{(n)} = [\max_{1 \le n \le N} y_{ij}^{(n)}]$ *is a matrix whose element occupying the ij-th entry is the largest of the N* $y_{ij}^{(n)}$ *elements.*

Notation 9 *i) The* κ_{ij}^{+} *and* κ^{+} *in Condition (C3a) in Theorem 1(B) below are obtained as follows: Let* $\kappa^{+} = \max[\mathbf{K}^{+}]$, *where* $\mathbf{K}^{+} = [\kappa_{ij}^{+}]$, *and* κ_{ij}^{+} *is the smallest integer greater than or equal to* $\max\{0, \varphi_{ij}^{+}\}$ *with* $\mathbf{\Phi}^{+} = [\varphi_{ij}^{+}]$, *given by*

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

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

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

and $\mathbf{A}^{+} = \sum_{l=1}^{q} \mathbf{A}^{(l)}$,
ii) Similarly, the κ_{ij}^{-} *and* κ^{-} *in Condition (C3b) below, are obtained as in (i) above by replacing the + superscript with the − one, where* $\mathbf{A}^{-}$ *now is given by* $\mathbf{A}^{-} = \sum_{l=1}^{q}(\mathbf{A}^{(l)} + \mathbf{\Gamma}^{(l)})$.

The following theorem holds.

4.1 Tractable Expressions

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

$$(B)\begin{cases} \phi_1 \text{ is real, and } \phi_1 > 0, & (C1) \\ adj[\mathbf{I}\phi_1 - \mathbf{B}]\mathbf{G}^{+} > \mathbf{0} & (C2a) \\ adj[\mathbf{I}\phi_1 - \mathbf{B}]\mathbf{G}^{-} > \mathbf{0} & (C2b) \\ \underset{(for\ each\ k^{+}=1,\dots,\kappa^{+})}{\mathbf{\Psi}^{+(k^{+})} \overset{(k^{+} \le \kappa_{ij}^{+})}{\ge} \mathbf{0}} & (C3a) \\ \underset{(for\ each\ k^{-}=1,\dots,\kappa^{-})}{\mathbf{\Psi}^{-(k^{-})} \overset{(k^{-} \le \kappa_{ij}^{-})}{\ge} \mathbf{0}} & (C3b) \end{cases}$$

(the symbol $\overset{(k^{+} \le \kappa_{ij}^{+})}{\ge}$ *means that if* $k^{+} > \kappa_{ij}^{+}$, *then the ij-th scalar inequality of the matrix inequalities in Condition (C3a), that is* $\mathbf{\Psi}^{+(k^{+})} \ge \mathbf{0}$, *should be disregarded).*

Theorem 1 follows directly from Proposition A1 in Section A of the online Appendix (see also Theorem 1 in Conrad and Karanasos, 2010) and Corollary 1.

Summary 1 *Interestingly, in the above Theorem we only have to check:*
i) from the condition in part (A) if all the N elements of the vector $adj[\mathbf{I} - \mathbf{B}]\boldsymbol{\omega}$ *are positive,*
ii) if all the $2N^2$ *elements of the two matrices in (C2) are positive, and*
iii) if all the N^2 *elements of each of the* $\kappa^{+} + \kappa^{-}$ *matrices in Conditions (C3) are non-negative (clearly, the latter condition, when* $k^{+} = k^{-} = 1$, *implies that* $\mathbf{A}^{(1)} \ge \mathbf{0}$ *and* $\mathbf{A}^{(1)} + \mathbf{\Gamma}^{(1)} \ge \mathbf{0}$*).*
In other words, we replace (for the symmetric case) $\alpha_{ij}(\phi_1^{-1}) > 0$, *for* $i, j = 1, \dots, N$, *in Proposition A1 in the online Appendix by its equivalent matrix expression* $adj[\mathbf{I}\phi_1 - \mathbf{B}]\mathbf{G}^{+} > \mathbf{0}$, *and likewise* $\psi_{ij}^{(k_{ij})} \ge 0$, *for* $i, j = 1, \dots, N$ *by its equivalent matrix expressions,* $\mathbf{\Psi}^{+(k^{+})} \ge \mathbf{0}$ *(see Conditions (C2a) and (C3a) in part (B) of the above Theorem, respectively).*

Next we present an important Remark (recall that $\mathbf{\Psi}_t^{(k)}$ has been defined in Definition 3, see also eq. (11) in Corollary 1).

Remark 4 *In practice one should just check the non-negativity constraints in Conditions (C3a) and (C3b) for all k^+ and k^- from 1 up to large enough κ^+, κ^-, respectively, i.e., $\kappa^+ = \kappa^- = Nq$. In other words, in practice the matrix inequalities in Conditions (C3) simplifies considerably since they reduce to: $\mathbf{\Psi}^{+(k)} \geq \mathbf{0}$ and $\mathbf{\Psi}^{-(k)} \geq \mathbf{0}$, for $k = 1, \ldots, \kappa$ with κ, for example, equal to Nq. 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.*

4.2 Scalar Inequalities

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.

Notation 10 *i) Let $\mathbf{B}^\star = [\beta_{ij}^\star]$ be given by $\mathbf{B}^\star = \mathbf{I} - \mathbf{B}$, and $\beta_{im}^{\star(a)} = (-1)^{i+m} \det[\mathbf{B}_{mi}^\star]$, where the $\mathbf{B}_{mi}^\star$ matrix is obtained by deleting the m-th row and the i-th column from $\mathbf{B}^\star$,*
ii) Similarly, let $\mathbf{B}^\triangleright = [\beta_{ij}^\triangleright]$ be given by $\mathbf{I}\phi_1 - \mathbf{B}$ and $\beta_{im}^{\triangleright(a)} = (-1)^{i+m} \det[\mathbf{B}_{mi}^\triangleright]$ where the latter matrix is obtained by deleting the m-th row and the i-th column from $\mathbf{B}^\triangleright$.

Then the condition in part (A) of Theorem 1 implies that:

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

Likewise, Conditions (C2) in Theorem 1 are equivalent to:

$$\begin{aligned} \sum_{m=1}^{N} \beta_{im}^{\triangleright(a)} g_{ij}^{+} &> 0, \text{ (C2a}'\text{)}, \\ \sum_{m=1}^{N} \beta_{im}^{\triangleright(a)} g_{ij}^{-} &> 0 \text{ (C2b}'\text{)}, \end{aligned}$$

for all $i, j = 1, \ldots, N$ (we recall that g_{ij}^+ and g_{ij}^- are given in Notation 6).

Notation 11 *i) Let $\boldsymbol{\phi} = [\phi_i]$ be the vector of the N distinct roots (see eq. (4)). Then there is a nonsingular matrix $\mathbf{\Lambda} = [\lambda_{ij}]$ (the matrix with the N eigenvectors of $\mathbf{B}$) such that*

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

ii) Denote the element occupying the ij-th entry of $\mathbf{\Lambda}^{-1}$ by λ_{ij}^*, that is, $\mathbf{\Lambda}^{-1} = [\lambda_{ij}^*]$.

Then, for each k, the simplified versions of the two Conditions (C3) in Theorem 1, (see also Remark 4), that is the ij-th entries of $\mathbf{\Psi}^{+(k)}$ and $\mathbf{\Psi}^{-(k)}$ must be non-negative for all $i, j = 1, \ldots, N$, amounts to:

$$\begin{aligned} \psi_{ij}^{+(k)} &= \sum\nolimits_{s=1}^{\min(q,k)} \sum\nolimits_{m=1}^{N} \sum\nolimits_{n=1}^{N} \lambda_{in} \lambda_{nm}^* \phi_n^{k-s} \alpha_{mj}^{(s)} > 0 \text{ (C3a}'\text{)}, \\ \psi_{ij}^{-(k)} &= \sum\nolimits_{s=1}^{\min(q,k)} \sum\nolimits_{m=1}^{N} \sum\nolimits_{n=1}^{N} \lambda_{in} \lambda_{nm}^* \phi_n^{k-s} \left(\alpha_{mj}^{(s)} + \gamma_{mj}^{(s)} \right) > 0 \text{ (C3b}'\text{)}. \end{aligned}$$

The above inequalities when $k = 1$ are: $\alpha_{ij}^{(1)} > 0$ and $\alpha_{ij}^{(1)} + \gamma_{ij}^{(1)} > 0$, since $\sum_{n=1}^{N} \lambda_{in} \lambda_{nm}^* = 1$ if $i = m$, and zero otherwise.

4.3 Trivariate System

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 symmetric case of order $(1,1)$. For notational simplicity we will use: $\mathbf{A} = \mathbf{A}^{(1)}$.

Notice that for the symmetric trivariate model Conditions C2a and C3a in Theorem 1B simplify to

$$adj[\mathbf{I}\phi_1 - \mathbf{B}]\mathbf{A} > \mathbf{0}, \text{ and } \mathbf{B}^{k-1}\mathbf{A} > \mathbf{0} \text{ for } k = 1, 2, 3.$$

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

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

$$(B)\begin{cases} \phi_1 \text{ is real, and } \phi_1 > 0, & (C1) \\ \sum_{m=1}^{3} \beta_{im}^{\triangleright(a)} \alpha_{mj} > 0, \text{ for all } i,j = 1,2,3, \text{ where} & \\ \beta_{im}^{\triangleright(a)} = \begin{cases} (\phi_1-\beta_{ll})(\phi_1-\beta_{nn}) - \beta_{ln}\beta_{nl} & \text{if } \ i = m,\ l \neq n \neq i, \\ (-1)^{i+m+1}[\beta_{im}(\phi_1-\beta_{ll}) - \beta_{il}\beta_{lm}] & \text{if } \quad i \neq m \neq l, \end{cases} & (C2a'') \\ a_{ij} \geq 0,\ \sum_{m=1}^{3} \beta_{im}\alpha_{mj} \geq 0, \text{ and } \sum_{m=1}^{3}\sum_{l=1}^{3} \beta_{il}\beta_{lm}\alpha_{mj} > 0, \text{ for all } i,j = 1,2,3. & (C3a'') \end{cases}$$

4.4 Numerical Examples

In what follows we graphically illustrate the necessary and sufficient parameter set for the trivariate UF 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 5 presented in Section 6 (in particular, the results from the FTSE100 index in dataset 2). The four data generating processes (DGP) are given by:

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

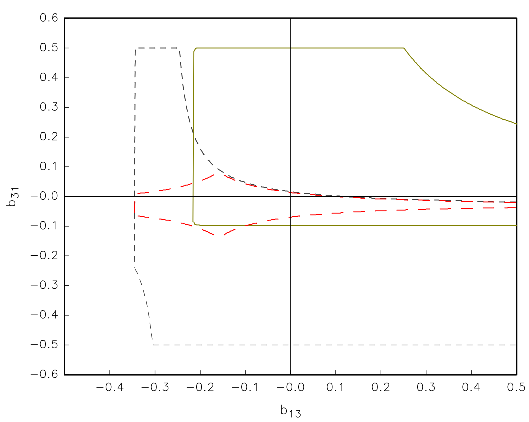
	DGP Ex.1	DGP Ex.2
$\boldsymbol{\omega}'$	$\begin{pmatrix} 0.149 & 0.074 & 0.124 \end{pmatrix}$	$\begin{pmatrix} 0.153 & 0.106 & 0.103 \end{pmatrix}$
$\mathbf{A}$	$\begin{pmatrix} 0.064 & 0.021 & 0.158 \\ 0.008 & 0.005 & 0.108 \\ 0.028 & 0.043 & 0.198 \end{pmatrix}$	$\begin{pmatrix} 0.075 & 0.017 & 0.123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

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

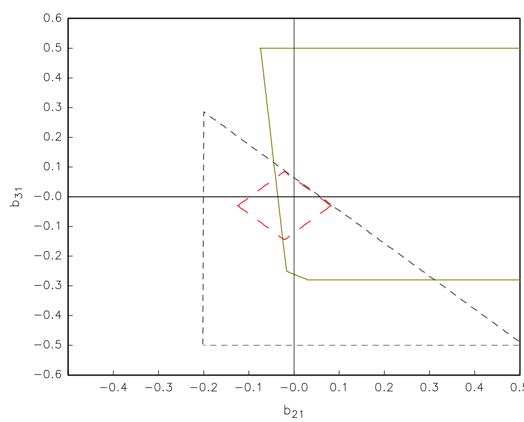
	DGP Ex.3			DGP Ex.4		
$\boldsymbol{\omega}'$	0.162	0.114	0.117	0.214	0.184	0.164
A	0.075	0.011	0.140	0.078	0.012	0.171
	0.013	0.003	0.139	0.012	0.005	0.100
	0.023	0.044	0.201	0.048	0.029	0.228
B	0.744	0.002	0.051	0.743	0.031	b_{13}
	b_{21}	0.901	b_{23}	$-\mathbf{0.028}$	0.851	0.053
	0.009	0.056	0.559	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 A [see Remark 5 and eq. (A.4)] and B [see eq. (B.2a) in Theorem 2 and Remark 7], 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.

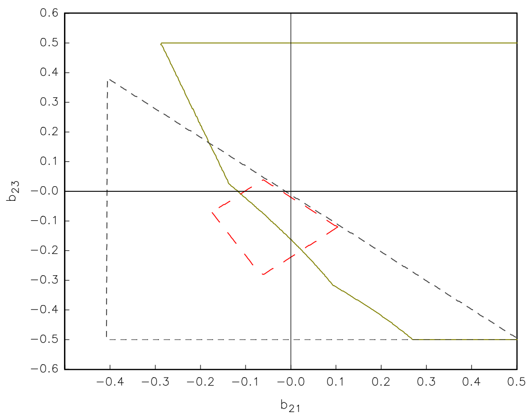
Example 2 is visualized in Figure 1b. The conditions of Theorem 1 allow for negative spillovers from σ_{1t} to σ_{2t} and σ_{3t}. The negative parameter set that satisfies all conditions simultaneously is given by the area that is above and to the right of all the three lines in the third quadrant. Figure 1c shows that, for the parameters in Example 3, the conditions of Theorem 1 allow for negative spillovers from σ_{1t} and σ_{3t} to σ_{2t}. 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.



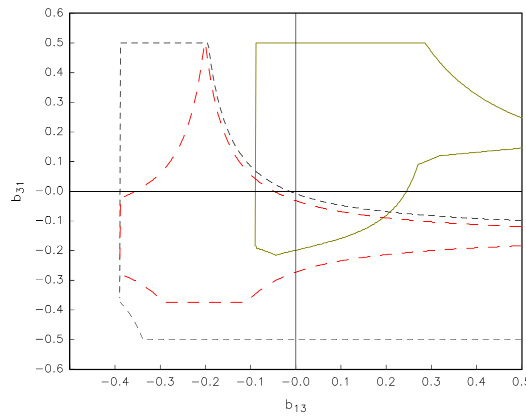
a. Example 1



b. Example 2



c. Example 3



d. Example 4

FIGURE 1. Necessary and sufficient parameter sets for the trivariate UF MEM$(1, 1)$ 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.

5 Monte Carlo Simulations

In this Section we employ the symmetric unrestricted full model of order $(1, 1)$, and by using Monte Carlo simulations, we examine the effects of ignoring the non-negativity conditions in Theorem 1 on: i) the bias of 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 (that is, that all parameters are allowed to take only non-negative values), 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 a conditional expectation and conditional covariance matrix $\mathbf{Q} = [q_{ij}]$ (using dataset 2, the FTSE100 index; see the empirical Section and Table 5 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
$\boldsymbol{\sigma}_{1t}$	0.142	0.314	1.017	4.572	4.534	14.887	9.903	6.667	48.354
$\boldsymbol{\sigma}_{2t}$	0.270	0.493	1.310	2.888	3.131	29.258	7.623	4.823	26.537
$\boldsymbol{\sigma}_{3t}$	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. (A.2) in Appendix A.

6 Empirical Results

6.1 The MUF MEM

In this Section we estimate trivariate and four variate unrestricted full MEM(1, 1) (that is symmetric models; see also Cipollini, et al. 2013):

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

where we recall that $\boldsymbol{\sigma}_t=[\sigma_{it}]$ and $\boldsymbol{\varepsilon}_t=\mathbf{Z}_t\boldsymbol{\sigma}_t$ with $\mathbf{Z}_t=diag[\mathbf{e}_t]$, $\mathbf{e}_t=[e_{it}]$ (see eq. (1)). Notice that $\boldsymbol{\varepsilon}_t$, is the vector, which contains the observed series. Recall that for the MEM, we assume that the stochastic vector $\mathbf{e}_t>\mathbf{0}_{N\times 1}$ (and, hence, $\boldsymbol{\varepsilon}_t>\mathbf{0}_{N\times 1}$) is $i.i.d$ with unit vector $\mathbf{j}$ as a conditional expectation (so $\boldsymbol{\sigma}_t=\mathbb{E}(\boldsymbol{\varepsilon}_t\,|\mathcal{F}_{t-1}))$, positive definite (time invariant) conditional correlation matrix $\mathbf{R}=[\rho_{ij}]$ with $\rho_{ii}=1$, and conditional covariance matrix $\mathbf{Q}=[q_{ij}]=\widetilde{\mathbf{Q}}^{-1/2}\mathbf{R}\widetilde{\mathbf{Q}}^{-1/2}$ (we recall that $\widetilde{\mathbf{Q}}$ denotes the $\mathbf{Q}$ matrix with its off-diagonal elements equal to zero, see also Section 2.2). The elements of $\boldsymbol{\varepsilon}_t$ 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 (conditional) distribution for the innovation vector $\mathbf{e}_t$, which is a random vector defined in $[\mathbf{0}_{Nx1},+\infty)$, that is $\mathbf{e}_t\,|\mathcal{F}_{t-1}\sim\ln N(\mathbf{j},\mathbf{Q})$. The log likelihood function, based on $\boldsymbol{\varepsilon}_t=\mathbf{Z}_t\boldsymbol{\sigma}_t$, is given by

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

where

$$\ln f(\varepsilon_t|\boldsymbol{\theta})=-\frac{N}{2}\ln(2\pi)-\frac{1}{2}\ln|\mathbf{Q}|-\sum\nolimits_{i=1}^{N}\ln\varepsilon_{it}-\frac{1}{2}(\ln\boldsymbol{\varepsilon}_t-\ln\boldsymbol{\sigma}_t-\boldsymbol{\xi})'\mathbf{Q}^{-1}(\ln\boldsymbol{\varepsilon}_t-\ln\boldsymbol{\sigma}_t-\boldsymbol{\xi}), \tag{13}$$

with $\boldsymbol{\xi}=-1/2\mathbf{q}$. These assume that $\boldsymbol{\theta}=[\boldsymbol{\beta}',\boldsymbol{\rho}]$, where $\boldsymbol{\beta}$ contains the parameters in $\boldsymbol{\sigma}_t$, 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 (see, for details, the supplementary Appendix A, which is available upon request).[8] In what follows we estimate three (two trivariate and one four-variate) UF 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.[9] 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).[10] 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 impose the matrix inequality constraints of Theorem 1. That is, combining restrictions on the parameter space with the sample data, we are using the constrained QML estimation.

Duration, Volume and Volatility

[8] 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.

[9] See Section B 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.

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

Manganelli (2005) used a diagonal **B** matrix and an unrestricted **A** matrix. It should be noted that Manganelli's results violate the non-negativity conditions in Theorem 1 (in Section D 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 **A** matrix is unrestricted. In our estimation all the parameters of the **A** matrix that turned out to be negative were constrained to zero (see the constrained QML estimation results in Table 4 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 constrained to zero. Similarly, all the parameters of the **B** matrix that turned out to be negative and violated the matrix inequality constraints were constrained 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 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 constrained 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 constrained 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 C1 in Section C of the supplementary Appendix).

Table 4. Trivariate UF MEM $(1,1)$ of intra day trading duration, stock volume and volatility.

	AVT			COX			CP			DLP			GAP		
A	**0.141**	-[1]	-	**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.
[1] The parameters took negative values, and, therefore, were constrained to zero.
The ω_i and the q_{ij}, $i, j = 1, \ldots, N$ are not reported but they are available upon request.

Table 5. Trivariate UF 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 5 above. However in the aforementioned paper, the estimated **A** matrix has negative elements and, therefore, the non-negativity conditions are violated (see Table D2 in Section D 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 C2 in Section C 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 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 6 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 C3 in Section C 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 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 Table D3 in Section D in the supplementary Appendix).

Table 6. Four variate UF 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}, $i, j = 1, \ldots, N$ are not reported but they are available upon request.

Table 7. Trivariate UFAP 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}, $i, j = 1, \ldots, N$ are not reported but they are available upon request.

6.2 The Asymmetric Power Specification

In this Section we estimate a trivariate UFA MEM$(1,1)$ with power (P) effects, as given by eq. **(12),** by replacing $\boldsymbol{\sigma}_t$ and $\boldsymbol{\varepsilon}_t$ with their power transformations: $\boldsymbol{\sigma}_t^{\wedge\boldsymbol{\delta}}$ and $|\varepsilon_t|^{\wedge\boldsymbol{\delta}}$, where $\boldsymbol{\delta} = [\delta_i]$ with $\delta_i(0,\infty)$ for all i. We use 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 constrained QML estimation, see eq. (13)). The first equation in Table 7 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 stock returns to the power transformed 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 Yfanti et al. (2021). 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 Conclusions

In this paper we have examined some of the properties of N-dimensional unrestricted full MEM. Our methodology can be applied to multivariate GARCH/HEAVY models as well. For the parameters of these systems we have derived matrix inequality constraints that require the conditional variables (means or variances) to be almost surely non-negative at all t. Our theroretical approach 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 papers report estimated parameters with negative values, which frequently violate the non-negativity conditions.

We have also 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. That is, we have employed the constrained QML estimation.

One 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 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.

Alternatively, one could employ a new mixture formulation (see, for example, **OUR NEW PAPER**), which might be an effective way to relax some of these constraints. In particular, one can use the exponential function in some but not all of the N equations. For example, in the multivariate GARCH/HEAVY model by replacing some of the conditional variances with their logarithms we can cut down the dimensions of the non-negativity conditions. 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. We leave this line of research for future work.

Our 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 systems 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 "one-sided" representation that we have proposed in this paper, coupled with the novel methodology, proposed in Karanasos et al. (2022), 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] Christodoulakis, G. A. (2007). "Common volatility and correlation clustering in asset returns." European Journal of Operational Research 182(3): 1263-1284.

[4] Christodoulakis, G. A. and S. E. Satchell (2002). "Correlated ARCH (CorrARCH): Modelling the time-varying conditional correlation between financial asset returns." European Journal of Operational Research 139(2): 351-370.

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

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

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

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

[9] 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.

[10] 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.

[11] 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.

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

[13] 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.

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

[15] 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.

[16] 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.

[17] 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.

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

[19] 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.

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

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

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

[23] Karanasos, M. (2000). "Some Exact Formulae for the Constant Correlation and Diagonal MGARCH Models." Discussion Paper, 2000/14, University of York.

[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., A. G. Paraskevopoulos, A. Magdalinos and A. Canepa (2022). "A unified theory for the ARMA models with time varying coefficients: One solution fits all." Working Paper, Brunel University, London.

[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] Ling, S. and M. McAleer (2003). "Asymptotic Theory for a Vector ARMA-GARCH Model." Econometric Theory 19(2): 280-310.

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

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

[31] 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.

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

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

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

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

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

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

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

[39] 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

[40] 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.

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

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

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

[44] S. Yfanti, G. Chortareas, M. Karanasos and E. Noikokyris (2022). "A three-dimensional asymmetric power HEAVY model." International Journal of Finance and Economics, forthcoming.

A APPENDIX

In this Appendix we will present an explicit formula for the k-step-ahead optimal (in L_2 sense) linear predictor of the MUFA MEM$(1,1)$:

$$(\mathbf{I}-\mathbf{B}L)\boldsymbol{\sigma}_t = \boldsymbol{\omega}+L\mathbf{A}_t\boldsymbol{\varepsilon}_t, \tag{A.1}$$

where $\mathbf{A}_t = \mathbf{A}_t^{(1)}$.

OPTIMAL PREDICTORS

First, we will introduce the following definitions.

Definition 4 *Define* $\mathbf{C}_t$ *as*

$$\mathbf{C}_t = \mathbf{B} + \mathbf{A}_t\mathbf{Z}_t,$$

with, $\mathbf{C}$ =$\mathbb{E}[\mathbf{C}_t]$*, given by*

$$\mathbf{C} = \mathbf{B} + \overline{\mathbf{A}},$$

where $\overline{\mathbf{A}} = \mathbb{E}[\mathbf{A}_t] = \mathbf{A} + \boldsymbol{\Gamma}\frac{1}{2}$*, and let* $\lambda_{\max}(\mathbf{C})$ *refer to the modulus of the largest eigenvalue of* $\mathbf{C}$.

(recall that: $\mathbb{E}[\mathbf{e}_t\,|\mathcal{F}_{t-1}] = \mathbf{j}$, $\mathbf{Z}_t = diag[\mathbf{e}_t]$, see eq. (1), and, thus $\mathbb{E}(\mathbf{Z}_t) = \mathbf{I}$).

Next, let $\mathcal{K}_{t-k-1}$, $k \in \mathbb{Z}_{\geq 1}$, be the smallest closed subspace of L_2 (see Condition 1 in the Appendix B) spanned by the sequence of the past observable random vectors $\{\mathbf{j}, \boldsymbol{\sigma}_{t-k}\}$. Following Karanasos et al. (2022) we shall denote the orthogonal projection of $\boldsymbol{\sigma}_t$ onto $\mathcal{K}_{t-k-1}$ by $\widehat{\mathbb{E}}(\boldsymbol{\sigma}_t\,|\mathcal{K}_{t-k-1})$ and we shall refer to it as the optimal linear predictor of $\boldsymbol{\sigma}_t$ based on $\mathcal{K}_{t-k-1}$.

Proposition 2 *The k-step-ahead optimal linear predictor of $\boldsymbol{\sigma}_t$ in eq. (A.1), based on $\mathcal{K}_{t-k-1}$, is given by*

$$\widehat{\mathbb{E}}(\boldsymbol{\sigma}_t\,|\mathcal{K}_{t-k-1}) = (\mathbf{I} - \mathbf{C})^{-1}(\mathbf{I} - \mathbf{C}^k)\boldsymbol{\omega} + \mathbf{C}^k\boldsymbol{\sigma}_{t-k}. \tag{A.2}$$

In addition, the first-order moment vector, $\boldsymbol{\sigma} = \mathbb{E}(\boldsymbol{\sigma}_t) = \lim_{k\to\infty} \widehat{\mathbb{E}}(\boldsymbol{\sigma}_t\,|\mathcal{K}_{t-k-1})$, exists in $\mathbb{R}_{>0}$ if and only if

$$\lambda_{\max}(\mathbf{C}) < 1. \tag{A.3}$$

Under the condition in (A.3), $\boldsymbol{\sigma}$ is given by

$$\boldsymbol{\sigma} = (\mathbf{I} - \mathbf{C})^{-1}\boldsymbol{\omega}. \tag{A.4}$$

Notice that, the k-step-ahead optimal linear predictor of $\boldsymbol{\varepsilon}_t$ (based on $\mathcal{K}_{t-k-1}$) is

$$\widehat{\mathbb{E}}(\boldsymbol{\varepsilon}_t\,|\mathcal{K}_{t-k-1}) = \widehat{\mathbb{E}}(\boldsymbol{\sigma}_t\,|\mathcal{K}_{t-k-1})],$$

and finally, under the condition in (A.3):

$$\boldsymbol{\varepsilon} = \mathbb{E}(\boldsymbol{\varepsilon}_t) = \lim_{k\to\infty} \widehat{\mathbb{E}}(\boldsymbol{\varepsilon}_t\,|\mathcal{K}_{t-k-1}) = \boldsymbol{\sigma}. \tag{A.5}$$

The proof of eq. (A.2) is presented in Section B.1 of the online Appendix.

Remark 5 *Notice that eq. (A.4) imposes, for the MEM, an additional matrix inequality constraint on the parameter space, that is $(\mathbf{I} - \mathbf{C})^{-1}\boldsymbol{\omega} > \mathbf{0}$.*

B APPENDIX

Now that we have presented the optimal predictors and the first unconditional moment for the MUFA MEM$(1, 1)$, we will examine its second moment structure.

SECOND MOMENTS

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

Notation 12 *i) Let*

$$\begin{aligned} \mathbb{E}\left(\mathbf{Z}_t^{\otimes 2}\right) &= \mathbb{E}\left(\mathbf{Z}_t \otimes \mathbf{Z}_t\right), \\ \widetilde{\mathbf{Z}} &= \mathbb{E}\left(\mathbf{Z}_t^{\otimes 2}\right) - \mathbf{I}_{N^2} = \mathbb{E}[(\mathbf{Z}_t - \mathbf{I}_N)^{\otimes 2}]. \end{aligned} \tag{B.1a}$$

ii) Let

$$\mathbf{C}^{\otimes 2} = \mathbf{C} \otimes \mathbf{C},\ \overline{\mathbf{A}}^{\otimes 2} = \mathbb{E}(\overline{\mathbf{A}} \otimes \overline{\mathbf{A}}). \tag{B.1b}$$

iii) Let

$$\widetilde{\mathbf{C}} = \mathbf{C}^{\otimes 2} + \overline{\mathbf{A}}^{\otimes 2}\ \widetilde{\mathbf{Z}}. \tag{B.1c}$$

(recall that $\mathbf{C}$ and $\overline{\mathbf{A}}$ have been defined in Definition 4).

Remark 6 *Notice that* $\widetilde{\mathbf{Z}}$ *in eq. (B.1a) is a diagonal matrix (of order* N^2*), and the element occupying its* r*-th diagonal entry, with* $r = [(i-1)N + j]$ *and* $i, j = 1, \ldots, N$*, is given by*

$$z_{i-1,N+j} = \mathbb{E}(e_{it}e_{jt}) - \mathbb{E}(e_{it})\mathbb{E}(e_{jt}) = q_{ij}.$$

Therefore, $\widetilde{\mathbf{Z}}\mathbf{j}_{N^2}$ *is a vector of order* N^2 *with the element occupying the* $[(i-1)N + j]$*-th entry given by* q_{ij}*.*

Definition 5 *i) Let* $\boldsymbol{\Gamma}(\ell) = [\gamma_{ij}(\ell)]$*,* $\ell \in \mathbb{Z}_{\geq 0}$*, denote the multidimensional covariance function of* $\{\boldsymbol{\sigma}_t\}$*, that is*

$$\boldsymbol{\Gamma}(\ell) = \mathbb{E}[(\boldsymbol{\sigma}_{t-\ell} - \boldsymbol{\sigma})(\boldsymbol{\sigma}_t - \boldsymbol{\sigma})'],$$

or

$$\boldsymbol{\Gamma}(\ell) = \boldsymbol{\Sigma}(\ell) - \boldsymbol{\sigma}\boldsymbol{\sigma}',$$

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

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

where $\gamma_{ii}(0)$ *is the element occupying the* i*-th diagonal entry of* $\boldsymbol{\Gamma}(0)$*. To further fix notation, write the* ℓ*-th-order, for* $\ell \geq 1$*, autocorrelation matrix of* $\{\boldsymbol{\sigma}_t\}$ *as*

$$\mathbf{R}(\ell) = \mathbf{D}^{-1}\boldsymbol{\Gamma}(\ell)\mathbf{D}^{-1}.$$

ii) Similarly to i), we will denote the multidimensional covariance function of $\{\boldsymbol{\varepsilon}_t\}$ *and its vec form, by* $\boldsymbol{\Gamma}_\varepsilon(\ell)$ *and* $\boldsymbol{\gamma}_\varepsilon(\ell)$*, respectively. That is, we will use the subscript* ε *to discriminate the covariance functions of* $\{\boldsymbol{\sigma}_t\}$ *and* $\{\boldsymbol{\varepsilon}_t\}$*.*

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

Theorem 2 *Consider the MUFA MEM(*$1,1$*) process. Under Condition 1 the vec form of* $\boldsymbol{\Gamma}(0)$ *exists in* $\mathbb{R}$*, and it is given by*

$$\boldsymbol{\gamma}(0) = \left(\mathbf{I}_{N^2} - \widetilde{\mathbf{C}}\right)^{-1} \overline{\mathbf{A}}^{\otimes 2}\widetilde{\mathbf{Z}}\boldsymbol{\sigma}^{\otimes 2}. \tag{B.2a}$$

Further, the vec form of the covariance function, that is $\boldsymbol{\gamma}(\ell)$*, for* $\ell \geq 1$*, is given by*

$$\boldsymbol{\gamma}(\ell) = \left(\mathbf{C}^\ell \otimes \mathbf{I}\right)\boldsymbol{\gamma}(0) \tag{B.2b}$$

(recall that $\widetilde{\mathbf{Z}}$*, and* $\widetilde{\mathbf{C}}$ *are given in eqs. (B.1a) and (B.1c), respectively).*

Remark 7 *Notice that eq. (B.2a) imposes an additional matrix inequality constraint on the parameter space, that is* $\mathbf{D}^2 > \mathbf{0}$*.*

Theorem 3 *Consider the* N*-dimensional vector MUFA MEM(*$1,1$*) process. Under Condition 1 the vec form of* $\boldsymbol{\Gamma}_\varepsilon(0)$*, exists in* $\mathbb{R}$*, and it is given by*

$$\boldsymbol{\gamma}_\varepsilon(0) \;=\; \mathbb{E}\left(\mathbf{Z}_t^{\otimes 2}\right)\boldsymbol{\gamma}(0) + \widetilde{\mathbf{Z}}\boldsymbol{\sigma}^{\otimes 2} = \left[\mathbb{E}\left(\mathbf{Z}_t^{\otimes 2}\right)\left(\mathbf{I}_{N^2} - \widetilde{\mathbf{C}}\right)^{-1}\overline{\mathbf{A}}^{\otimes 2} + \mathbf{I}_{N^2}\right]\widetilde{\mathbf{Z}}\boldsymbol{\sigma}^{\otimes 2}. \tag{B.3a}$$

Similarly, $\boldsymbol{\gamma}_\varepsilon(\ell)$*, for* $\ell \geq 1$*, is given by*

$$\begin{aligned}
\boldsymbol{\gamma}_\varepsilon(\ell) &= \boldsymbol{\gamma}(\ell) + (\mathbf{C}^{\ell-1} \otimes \mathbf{I})(\overline{\mathbf{A}} \otimes \mathbf{I})\widetilde{\mathbf{Z}}[\boldsymbol{\gamma}(0) + \boldsymbol{\sigma}^{\otimes 2}] \\
&= (\mathbf{C}^{\ell-1} \otimes \mathbf{I})\left\{[(\mathbf{C} \otimes \mathbf{I}) + (\overline{\mathbf{A}} \otimes \mathbf{I})\widetilde{\mathbf{Z}}]\left(\mathbf{I}_{N^2} - \widetilde{\mathbf{C}}\right)^{-1}\overline{\mathbf{A}}^{\otimes 2} + (\overline{\mathbf{A}} \otimes \mathbf{I})\right\}\widetilde{\mathbf{Z}}\boldsymbol{\sigma}^{\otimes 2}.
\end{aligned}$$

The proofs are presented in Section B.2 of the online Appendix.