EXPLICIT MODELING OF TIME VARIATION
OF DIEBOLD AND YILMAZ
CONNECTEDNESS MEASURES

By

NICOLA PIAZZALUNGA

A dissertation submitted in partial fulfillment

Of the requirements for the degree of

Doctor of Philosophy

[EDHEC Business School]

May 2018

Dissertation Committee

Abraham Lioui, PHD

Kamil Yilmaz, PHD

René Garcia, PHD

EDHEC BUSINESS SCHOOL

APPROVAL FORM

CANDIDATE'S NAME: Nicola Piazzalunga

TITLE OF DISSERTATION: **"Explicit Modeling of Time Variation of Diebold and Yilmaz Connectedness Measures"**

APPROVAL:

Abraham Lioui		*Chair*
Kamil Yilmaz		*External Examiner*
René Garcia		*Member*

☐ with minor revisions

☐ with major revisions

DATE: May 23, 2018

Abstract

Connectedness is a concept which is central to many areas of scientific interest. For what regards Economics and Finance, a remarkable example of the importance of the analysis of connectedness is the economic and financial crisis that begun around 2007.

One of the issues debated about that phenomenon is, in fact, the exceptional amount of connectedness characterizing a vast set of features of the financial sector: market risk, portfolio management, pricing and systematic risk are all examples of areas of study and practice that have been moved by some sort of connectedness among them and inside their categories.

One other way to state the problem is then to consider the mentioned crisis a highly multidimensional phenomenon, as connectedness arise as a relationship between a set of variables, and treat its complexity as a non linear process.

To proceed forward, the first question to answer would be how to model the concept of connectedness. The Diebold and Yilmaz framework adopted in this dissertation offers a methodology intended to identify connectedness and in a way that accommodates both its multidimensional nature and the non linearity requirement.

The multidimensional issue is solved by the Diebold and Yilmaz framework by identifying the connectedness measures through forecast error variance decompositions of vector autoregressions (VARs). Such a formulation is also key for framing connectedness as a feature influenced both by contemporaneous and lagged relationships among variables.

More specifically, connectedness measurements are identified as "forecast error variation in various locations (firms, markets, countries, etc.) due to shocks arising elsewhere". In practice, forecast error variance decompositions return a square matrix of forecast error variance shares of the order the number of system variables. Each variance share is either directed from a system variable to another, or from the system variable to itself. Variance shares are, moreover, the primitives

of aggregates that describe more complex relationships, as, for instance, the total connectedness in the system, the amount of connectedness directed from the whole system to one of its variables and vice versa, or even connectedness directed from one subset of the system variables to another subset of them.

The forecast error variance decomposition framework requires an identification method. The classic way to do it is to declare a causal order among the system variables and identify the forecast error variance decompositions through the orthogonal decomposition of the variance covariance matrix of the errors. Another way is the generalised (with respect to the order of the system variables) method, which needs a tractable distributional assumption of the (conditional) error term, typically normality. In multidimensional and non linear settings, the second choice might be preferable, as the causal chain can be an ambiguous structure to determine.

What about the non linearity issue? One easy way to accommodate for nonlinearity that has been shown to be a very general approximation of nonlinear models, which parameters change smoothly, or which parameters abrupt change can be predicted by a "hidden" linear model, is to formulate the nonlinearity as a linear model with time varying parameters. In practice, one estimates a linear VAR using the rolling windows technique, allowing thus the connectedness measures to vary over time, according to the characteristics of the sample in each window.

Another way to deal with nonlinearity is to explicitly model it, allowing thus the analyst to make a priori decisions about the form of the system distribution. This dissertation offers novel techniques that enable to explicitly model connectedness measures for both GARCH-DCC (first chapter) and Markov Switching (second chapter) specifications of VARs and it does so by extending the variance decomposition framework, both classic and generalized.

Contents

This thesis has been made possible by many contributions. I would like to thank my parents, Claudia Piatti and Roberto Piazzalunga, for their financial support and for resisting with me in this long and harduous journey. I would like to thank my supervisor, Professor René Garcia, for his patience and help, and Professor Abraham Lioui for his leadership and coaching during the final stage of the development of this thesis. I thank EDHEC Risk for the financial support during my first PhD year and its faculty members for their leadership and teachings. In particular, I would like to thank Professor Stéphane Gregoir for his teachings during my first PhD year in Paris. I would also like to thank Maud Gauchon and Brigitte Bogaerts-Chevillotte and all the administrative staff for helping in sorting out the miriads of necessities I had as a student. I would finally like to thank all the fellow PhD students of the EDHEC PhD in Finance Programme who participated with me in long and exceptionally interesting debates and exchange of ideas. I am honoured to have been part of such a vibrant and brilliant community.

List of Tables

List of Figures

Chapter 1

Assessment of the Impact of Multivariate Heteroscedasticity on the Dynamics of the Diebold and Yilmaz Measures of Connectedness. The Realized Volatility Case.

Nicola Piazzalunga

September 13, 2018

Assessment of the Impact of Multivariate Heteroscedasticity on the Dynamics of the Diebold and Yilmaz Measures of Connectedness.

The Realized Volatility Case.

Abstract

This paper introduces multivariate heteroscedastic effects in the dynamics of measures of connectedness proposed and developed by Diebold and Yilmaz (Diebold and Yilmaz [8], Diebold and Yilmaz [9], Diebold and Yılmaz [10], Diebold and Yilmaz [11]). The measures are based on forecast error variance decompositions. We consider variability in the covariance matrix of the error vector by using the Dynamic Conditional Correlation (DCC) model of Engle [12]. We develop this framework both for the Orthogonal Variance Decomposition (OVD) and Generalized Variance Decomposition (GVD) measures of connectedness. We apply this new methodology to measure the variability in the links between a set of weekly time series of realized variances for returns of equity indices. We confirm the presence of heteroscedastic effects and compare our results to a rolling-window methodology.

1.1 Introduction

As developed in Diebold and Yilmaz [11], issues of connectedness arise in many areas of finance and economics. Financial crises and contagion have of course surfaced as a fruitful area of application in the last ten years or so, but the concept of capturing an aggregate effect from individual firm or consumer network links is pervasive in economics. The main challenge when one wants to gauge systemic effects based on individual actions is to have a reliable measure. Connectedness provides such a methodology anchored in modern network theory.

Let us consider systemic risk. One way to think about systemic risk is in terms of connectedness among financial variables. Large comovements in financial asset markets or in the balance sheets of financial institutions depend on the strength of the connections between these asset markets, that is their connectedness. Diebold and Yilmaz [8] introduced a framework to measure connectedness among a set of financial variables based on a vector autoregression (VAR) forecast error variance decompositions. These are measures of the shares of the forecast error variance attributable to the shocks in individual components, leading to a matrix of dimension the number of system variables.

The variance decompositions normally used in the literature, the Orthogonal Variance Decompositions (OVD), are based on a specific a priori causal chain, or Wold Order (WO) (Wold [24]), and obtained through a Cholesky factorization. The order of the decomposition is often difficult to establish. Which variable comes first in the chain of transmissions? For this reason, the framework was extended by Diebold and Yilmaz [9] to allow researchers to be agnostic about the order of the decomposition. They propose variance decompositions which are invariant to ordering called Generalized Variance Decompositions (GVD). For this decomposition, shocks do not need to be orthogonal but additional assumptions are needed, typically normality.

The framework has been extensively applied to networks of financial variables, being them returns or volatilities, as, just for example, in Diebold and Yılmaz [10], Bostanci and Yilmaz [5], Demirer et al. [6] and Diebold et al. [7].

To put forward our contribution, let us write the simplest framework with orthogonal shocks:

$$y_t = \nu + \sum_{l=1}^{p} A_l y_{t-l} + u_t. \tag{1.1}$$

Given a particular sorting of the system variables, the Cholesky lower triangular decomposition of the conditional variance covariance matrix of the errors is:

$$\Sigma = PP'. \tag{1.2}$$

Based on this system, static measures of connectedness can be computed. However, as stressed by Diebold and Yilmaz [11], connectedness may well be dynamic in many applications. They suggest that the A_l coefficients may be time-varying (following random walks, for example) or that a rolling window is used for estimation. The latter approach is simple but requires the length of the window to be chosen appropriately, at the risk of producing a measure that is either too smooth or too choppy. By adopting such a method, not only the parameters A_l will vary with time, but the covariance matrix of the errors Σ will also change with each window. In this paper, we propose to adopt a specific data generating process for capturing the time variation of the covariance matrix of the error terms. We adopt the Dynamic Conditional Correlation (DCC) of Engle [12] with a diagonal scalar specification.

This property has relevance, as variances and correlations can vary between different states of the world, in particular, during financial crises. Forbes and Rigobon [16] make a distinction between interdependence and contagion based precisely on the fact that the strong links (measured by correlations) between markets remain strong in all states of the world. It is therefore important to allow for shocks of different amplitudes to obtain an unbiased measure of contagion.

To illustrate this modeling strategy, we measure the time-varying connectedness of a panel of realized variances of returns for several international indices. Our results indicate that the explicit modeling of heteroscedasticity matters indeed and can potentially significantly impact the connectedness measures, especially during crises.

Other studies employed a GARCH-DCC framework in a Diebold and Yilmaz (DY) connectedness analysis, but none, to our knowledge, has yet extended the variance decomposition technique to explicitly and formally embed heteroscedasticity into the variance decompositions. For example, Antonakakis [2] uses a GARCH-DCC model to estimate conditional volatilities to be used as system variables in a standard VAR, as Antonakakis et al. [3] do. In both cases, the dynamics of the connectedness measures were obtained through rolling windows estimations of linear VARs.

The remainder of the paper proceeds as follows. Section 2 introduces the two main measures of connectedness introduced by Diebold and Yilmaz (OVD and GVD), while Section 3 develops the diagonal scalar DCC model. In Section 4, we present the extension of the Diebold and Yilmaz framework to explicitly account for heteroscedastic effects. Section 5 provides an empirical illustration of the concept with realized variances of returns for several international indices. We conclude in Section 6.

1.2 The Diebold and Yilmaz Framework

As mentioned in the introduction, there are basically two methodological approaches to measuring connectedness. The first relies on a variance decomposition that is based on orthogonalized residuals from a given ordering on the vector autoregressive (VAR) system of variables under study. This is the OVD measure of connectedness that we present first. The second decomposition is not dependent on a specific WO and is invariant to ordering, hence its denomination Generalized Variance Decomposition (GVD), but it requires a tractable distributional assumption about the residuals of the VAR system, typically normality. Diebold and Yılmaz [10] report that total connectedness is robust to Cholesky ordering, but that directional connectedness, from one element to the others or to one element from the others, is more sensitive to this ordering. Therefore the GVD connectedness measures appear useful, especially for large systems or when it is not possible or desirable to impose a specific WO.

1.2.1 OVD Based Measures of Connectedness

Consider a $(K \times 1)$ vector $y_t = [y_{1t},\, y_{2t},\, \dots,\, y_{Kt}]$ and the following $VAR\,(p)$:

$$y_t = \nu + \sum_{l=1}^{p} A_l y_{t-l} + u_t, \tag{1.3}$$

where A_l are $(K \times K)$ matrices of autoregressive parameters and u_t is an i.i.d. normally distributed error vector with mean zero and variance covariance matrix $\Sigma = [\sigma_{ij}]$.

The process can be rewritten as

$$A\,(L)\, y_t = \nu + u_t, \tag{1.4}$$

where $A\,(L) = (I_K - A_1 L - \ldots - A_p L^p)$ and L is the lag operator.

Assuming stationarity, the system has the following infinite MA representation:

$$y_t = \mu + \Phi\,(L)\, u_t, \tag{1.5}$$

where $\Phi\,(L) = (I_K + \Phi_1 L + \Phi_2 L^2 + \ldots) : \ \Phi\,(L)\, A\,(L) = I_K$, $\mu = \nu / A\,(L)$ and Φ_n are $(K \times K)$ matrices of moving average parameters such that $\Phi_0 = I_K$ and

$$(\Phi_h |\, h > 0) = \sum_{l=1}^{\min(h,p)} A_l\, \Phi_{h-l}. \tag{1.6}$$

Following Lütkepohl [21], it is possible to define $y_t\,(H)$, the H-steps ahead forecast conditional on the information present in t, as

$$y_t\left(H\right) = \mu + \sum_{h=H}^{\infty} \Phi_h u_{t+H-h}. \tag{1.7}$$

As 1.3 can be rewritten as

$$y_t = \mu + \sum_{h=0}^{\infty} \Phi_h u_{t-h}, \tag{1.8}$$

the H-steps ahead forecast error conditional on the information in t is given by:

$$\begin{aligned}
\xi_t\left(H\right) &= y_{t+H} - y_t\left(H\right) \\
&= \sum_{h=0}^{\infty} \Phi_h u_{t+H-h} - \sum_{h=H}^{\infty} \Phi_h u_{t+H-h} \\
&= \sum_{h=0}^{H-1} \Phi_h u_{t+H-h}.
\end{aligned} \tag{1.9}$$

Finally, the MSE of the forecast can be computed in the following way:

$$\begin{aligned}
MSE\left[y_t\left(H\right)\right] &= E_t\left[\xi_t\left(H\right)\xi_t'\left(H\right)\right] \\
&= E_t\left[\sum_{h=0}^{H-1} \Phi_h u_{t+H-h} u_{t+H-h}' \Phi_h'\right] \\
&= \sum_{h=0}^{H-1} E_t\left[\Phi_h u_{t+H-h} u_{t+H-h}' \Phi_h'\right] \\
&= \sum_{h=0}^{H-1} \Phi_h E_t\left[u_{t+H-h} u_{t+H-h}'\right] \Phi_h'.
\end{aligned} \tag{1.10}$$

Assuming homoscedasticity, the MSE will be

$$MSE\left[y_t\left(H\right)\right] = \sum_{h=0}^{H-1} \Phi_h \Sigma \Phi_h'. \tag{1.11}$$

If the analyst is able to produce an a priori WO for the variables of the system, the Diebold and Yilmaz measures of connectedness could be computed by making the residuals orthogonal via a Cholesky lower triangular decomposition of the variance covariance matrix. Although the procedure is well-known, we recall the main steps of the algorithm in Appendix 1.A.

For every sorting of the system variables, the Cholesky lower triangular decomposition of the variance covariance matrix is:

$$\Sigma = PP'. \tag{1.12}$$

It follows that, if we define

$$w_t = P^{-1}u_t, \tag{1.13}$$

w_t is an error term with mean zero and

$$E\left(w_t w_t'\right) = \Sigma_w = E\left(P^{-1}u_t u_t'\left(P^{-1}\right)'\right) = P^{-1}\Sigma\left(P^{-1}\right)' = I_K. \tag{1.14}$$

As their variance covariance matrix is the identity matrix, the elements of w_t are, of course, orthogonal. The MA process can then be rewritten as

$$y_t = \mu + \mathit{\Theta}\left(L\right)w_t, \tag{1.15}$$

where $\Theta\left(L\right) = \left(\Theta_0 + \Theta_1 L + \Theta_2 L^2 + \ldots\right) = \Phi\left(L\right)P$.

The H-steps ahead forecast will be equal to:

$$y_t\left(H\right) = \mu + \sum_{h=H}^{\infty} \mathit{\Phi}_h u_{t+H-h}$$

$$
\begin{aligned}
&= \mu + \sum_{h=H}^{\infty} \Phi_h P P^{-1} u_{t+H-h} \\
&= \mu + \sum_{h=H}^{\infty} \Theta_h w_{t+H-h}, \qquad (1.16)
\end{aligned}
$$

with forecast error

$$
\begin{aligned}
\xi_t(H) &= y_{t+h} - y_t(H) \\
&= \sum_{h=0}^{\infty} \Theta_h w_{t+H-h} - \sum_{h=H}^{\infty} \Theta_h w_{t+H-h} \\
&= \sum_{h=0}^{H-1} \Theta_h w_{t+H-h}. \qquad (1.17)
\end{aligned}
$$

As the elements of w are not correlated, the mean squared error is

$$
\begin{aligned}
\text{MSE}\left[y_t(H)\right] &= E_t\left[\xi_t(H)\,\xi_t'(H)\right] \\
&= \sum_{h=0}^{H-1} \Theta_h \Theta_h'. \qquad (1.18)
\end{aligned}
$$

The diagonal elements of the MSE are the mean squared errors of each system equation and, for the i-th equation, can be expressed as:

$$
\begin{aligned}
e_i' \text{MSE}\left[y_t(H)\right] e_i &= \sum_{h=0}^{H-1} e_i' \Theta_h \Theta_h' e_i \\
&= \sum_{h=0}^{H-1} \sum_{j=1}^{K} \theta_{h,ij}^2. \qquad (1.19)
\end{aligned}
$$

The contribution to the MSE of the i-th equation coming from the j-th variable are thus

$$\begin{aligned}\sum_{h=0}^{H-1}\theta_{h,\,ij}^{2} &= \textstyle\sum_{h=0}^{H-1} e_i'\mathit{\Theta}_h e_j e_j'\mathit{\Theta}_h' e_i \\ &= \textstyle\sum_{h=0}^{H-1}\left(e_i'\mathit{\Theta}_h e_j\right)^2 .\end{aligned}$$

Then, the generic element of the $(K \times K)$ upper left quadrant of the connectedness table is a index based on the (scaled) orthogonal variance decomposition:

$$\begin{aligned}d_{t,\,ij}^{(OVD)}(H) &= \frac{\sum_{h=0}^{H-1}\left(e_i'\Theta_h e_j\right)^2}{\sum_{h=0}^{H-1} e_i'\Theta_h\Theta_h' e_i} \\ &= \frac{\sum_{h=0}^{H-1}\theta_{h,\,ij}^{2}}{\sum_{h=0}^{H-1}\sum_{j=1}^{K}\theta_{h,\,ij}^{2}} \qquad (1.20)\end{aligned}$$

so that the Orthogonal Measure of Total Connectedness (OMTC) can be computed as

$$\begin{aligned}S_t^{(OVD)}(H) &= \frac{\sum_{i,j=1,\,i\neq j}^{K} d_{t,\,ij}^{(OVD)}(H)}{\sum_{i,j=1}^{K} d_{ij}^{(OVD)}(H)} \\ &= \frac{1}{K}\sum_{i,j=1,\,i\neq j}^{K} d_{t,\,ij}^{(OVD)}(H). \qquad (1.21)\end{aligned}$$

1.2.2 GVD Based Measures of Connectedness

In this section, we briefly report the structure of generalized variance decompositions introduced Pesaran and Shin [23] and Koop et al. [20]. Consider again a $(K \times 1)$ vector $y_t = [y_{1t},\, y_{2t},\, \ldots,\, y_{Nt}]$ and the following $VAR\,(p)$:

$$y_t = \nu + \sum_{l=1}^{p} A_l y_{t-l} + u_t. \qquad (1.22)$$

Conditional on the information in t and on shocks to the i-th equation up to time $t + H$, that is, $u_{it},\ u_{i,\,t+1},\ \ldots,\ u_{i,\,t+H}$, the H-steps ahead forecast are equal to:

$$y_t^{(i)}\left(H\right) = \mu + \sum_{h=H}^{\infty} \varPhi_h u_{t+H-h} + \sum_{h=0}^{H-1} E_t\left(u_{t+H-h} \middle| u_{i,\,t+H-h}\right). \tag{1.23}$$

The conditional expectation in 1.23 depends on distributional assumptions regarding u_t. Assuming that $u_t \sim N\left(0,\,\Sigma\right)$,

$$E_t\left(u_{t+H-h} \middle| u_{i,\,t+H-h}\right) = \left(\sigma_{ii}^{-1}\Sigma e_i\right) u_{i,\,t+H-1-h}, \tag{1.24}$$

where e_i is the i-th column of the identity matrix of order K, Then, the H steps ahead forecast of y_t, conditional on given future shocks to the i-th equation, can be represented as

$$\begin{aligned} y_t^{(i)}\left(H\right) &= \mu + \sum_{h=H}^{\infty} \varPhi_h u_{t+H-h} + \sum_{h=0}^{H-1} \varPhi_h E_t\left(u_{t+H-h} \middle| u_{i,\,t+H-h}\right) \\ &= \mu + \sum_{h=H}^{\infty} \varPhi_h u_{t+H-h} + \sum_{h=0}^{H-1} \varPhi_h \left(\sigma_{ii}^{-1}\Sigma e_i\right) u_{i,\,t+H-h}, \end{aligned} \tag{1.25}$$

where the superscript (i) indicates it is conditional to shocks to the i-th equation. Then, the conditional H-steps ahead forecast errors can be computed as

$$\begin{aligned} \xi_t^{(i)}\left(H\right) &= y_{t+h} - y_t^{(i)}\left(H\right) \\ &= \sum_{h=0}^{H-1} \varPhi_h u_{t+H-h} - \sum_{h=0}^{H-1} \varPhi_h \left(\sigma_{ii}^{-1}\Sigma e_i\right) u_{i,\,t+H-h} \\ &= \sum_{h=0}^{H-1} \varPhi_h \left[u_{t+H-h} - \left(\sigma_{ii}^{-1}\Sigma e_i\right) u_{i,\,t+H-h}\right]. \end{aligned} \tag{1.26}$$

In Appendix 1.B it is showed that

$$
\begin{aligned}
MSE\left[y_t^{(i)}(H)\right] &= E_t\left[\xi_t^{(i)}(H)\,\xi_t^{(i)\prime}(H)\right] \\
&= \sum_{h=0}^{H-1} \Phi_h \Sigma \Phi_n' - \sigma_{ii}^{-1} \sum_{h=0}^{H-1} \left(\Phi_h \Sigma e_i e_i' \Sigma \Phi_h'\right) \\
&= \sum_{h=0}^{H-1} \Phi_h \Sigma \Phi_h' - \sigma_{ii}^{-1} \sum_{h=0}^{H-1} \left(\Phi_h \Sigma e_i\right)^2, \qquad (1.27)
\end{aligned}
$$

so that it is possible to define a quantity $\Delta_{t,\,i}(H)$ such that

$$
\begin{aligned}
\Delta_{t,\,i}(H) &= MSE\left[\xi_t(H)\right] - MSE\left[\xi_t^{(i)}(H)\right] \\
&= \sigma_{ii}^{-1}\left(\sum_{h=0}^{H-1} \Phi_h \Sigma e_i\right)^2. \qquad (1.28)
\end{aligned}
$$

The second term appearing in 1.28 measures the information brought by conditioning on the path of u_{it} and the reduction in the total mean-squared error. Every $(j,\,j)$ element of $\Delta_{t,\,i}(H)$ is a component of the the forecast error variance of the j-th equation accounting for the ignorance about the i-th equation and is therefore the amount of variance coming from variable i and directed to each of the $j = 1,\,2,\,\ldots,\,K$ variables. Writing this Δ expression for each variable in the system constitutes a variance decomposition. These quantities are called Generalized Variance Decompositions (GVDs), because they do not assume a specific WO.

The original formulation of the GVD, included in Pesaran and Pesaran [22], scales the $(j,\,j)$ elements of $\Delta_{t,\,i}(H)$ by the total MSE in the j-th equation:

$$
d_{t,\,ij}^{(GVD)}(H) = \frac{\sigma_{ii}^{-1}\sum_{h=0}^{H-1}\left(e_j' \Phi_h \Sigma e_i\right)^2}{\sum_{h=0}^{H} e_j' \Phi_h \Sigma \Phi_h' e_j}. \qquad (1.29)
$$

Since the GVD includes effects due to covariances (unless the variance covariance matrix of u is diagonal), forecast error variance contributions do not sum to one, because such sum is not the MSE of the i-th variable. To standardize the measure, a normalization can be adopted that depends on the focus of the analysis. For instance, if the analyst wanted a index for the variance spread from each variable through the rest of the system, she could use the following normalization:

$$\tilde{d}_{t,\,ij}^{(GVD)}\left(H\right)=\frac{d_{t,\,ij}^{(GVD)}\left(H\right)}{\sum_{i=1}^{K}d_{t,\,ij}^{(GVD)}\left(H\right)}. \tag{1.30}$$

By construction then, $\sum_{i=1}^{K}\tilde{d}_{t,\,ij}^{(GVD)}\left(H\right)=1$ and $\sum_{i,\,j=1}^{K}\tilde{d}_{t,\,ij}^{(GVD)}\left(H\right)=N$. Notice also that, with such normalization, $\sum_{j=1}^{K}\tilde{d}_{t,\,ij}^{(GVD)}\left(H\right)\neq 1$.

To obtain a index of the variance received from other variables, then the normalizations to use will be:

$$\tilde{d}_{t,\,ij}^{(g)}\left(H\right)=\frac{d_{t,\,ij}^{(g)}\left(H\right)}{\sum_{j=1}^{K}d_{t,\,ij}^{(g)}\left(H\right)}. \tag{1.31}$$

The Generalized Measure of Total Connectedness (GMTC) can then be computed as

$$\begin{aligned} S_{t}^{(GVD)}\left(H\right) &= \frac{\sum_{i,\,j=1,\,i\neq j}^{K}\tilde{d}_{t,\,ij}^{(GVD)}\left(H\right)}{\sum_{i,\,j=1}^{K}\tilde{d}_{t,\,ij}^{(GVD)}\left(H\right)} \\ &= \frac{1}{K}\sum_{i,\,j=1,\,i\neq j}^{K}\tilde{d}_{t,\,ij}^{(GVD)}\left(H\right), \end{aligned} \tag{1.32}$$

which returns the same number independently from the normalization scheme chosen.

1.2.3 The Connectedness Table

Diebold and Yilmaz [8] and Diebold and Yilmaz [9] introduced and extended the connectedness table as a $(K+1)\times(K+1)$ matrix containing both variance decompositions and aggregate measures. The latter are distinguished in three main blocks. The first two blocks of aggregate measures contain K total directional connectedness measures coming from other variables and K total directional connectedness measures going to other variables. The third main block is simply the total variance spilled across the system.

Orthogonal

The orthogonal connectedness table contains in the upper left quadrant a $(K \times K)$ orthogonal variance decompositions. The upper right block is made by a $(K \times 1)$ vector of total directional connectedness measures coming from other variables, while the lower left block is made by a $(1 \times K)$ vector of total directional connectedness going to other variables. Finally, the bottom right block contains a scalar, which is the total connectedness.

$$\begin{array}{cccccc}
 & y_1 & y_2 & \cdots & y_K & From\, Others \\
 & & & & & \\
y_1 & d_{11}^H & d_{12}^H & \cdots & d_{1K}^H & \sum_{j=1}^K d_{1j}^H,\, j \neq 1 \\
y_2 & d_{21}^H & d_{22}^H & \cdots & d_{2K}^H & \sum_{j=1}^K d_{2j}^H,\, j \neq 2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
y_K & d_{K1}^H & d_{K2}^H & \cdots & d_{KK}^H & \sum_{j=1}^K d_{Kj}^H,\, j \neq K \\
 & & & & & Total \\
 & \sum_{i=1}^K d_{i1}^H, & \sum_{i=1}^K d_{i1}^H,\, i \neq 1 & \cdots & \sum_{i=1}^K d_{i1}^H,\, i \neq 1 & \sum_{i,\,j=1}^K d_{ij}^H \\
To\, Others & i \neq 1 & i \neq 2 & & i \neq K & i \neq j
\end{array} \tag{1.33}$$

Generalized

The generalized connectedness table is similar to the orthogonal version, with a few exceptions. Firstly, the upper left quadrant contains generalized variance decompositions. Secondly, this table has an "opposite" logic with respect to the orthogonal version, in the sense that directional totals are transposed, with total to others in the upper right block and total from others in the lower left block. Such a form is due to the different construction of the generalized variance decompositions.

$$\begin{array}{cccccc} & y_1 & y_2 & \cdots & y_K & To\,Others \\ \\ y_1 & d_{11}^H & d_{12}^H & \cdots & d_{1K}^H & \sum_{j=1}^K d_{1j}^H,\, j \neq 1 \\ y_2 & d_{21}^H & d_{22}^H & \cdots & d_{2K}^H & \sum_{j=1}^K d_{2j}^H,\, j \neq 2 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ y_K & d_{K1}^H & d_{K2}^H & \cdots & d_{KK}^H & \sum_{j=1}^K d_{Kj}^H,\, j \neq K \\ & & & & & Total \\ From\,Others & \sum_{i=1}^K d_{i1}^H, & \sum_{i=1}^K d_{i1}^H,\, i \neq 1 & \cdots & \sum_{i=1}^K d_{i1}^H,\, i \neq 1 & \sum_{i,\,j=1}^K d_{ij}^H \\ & i \neq 1 & i \neq 2 & & i \neq K & i \neq j \end{array} \tag{1.34}$$

Building Specialized Aggregate Measures of Connectedness

The mathematical structure of variance decompositions allows the analyst to customize the way aggregate measures are computed. For example, one can select subsets of the variance decompositions table or build net spillovers to identify major contributors to the variance of the system. The interested reader may refer to Diebold and Yilmaz [9] for a more detailed exposition of the possible combinations.

1.3 The Diagonal Scalar DCC Model

1.3.1 The GARCH(1, 1) Diagonal Scalar DCC Model

Following Engle [13], the DCC framework is specified as follows.

Consider a vector $u_t = [u_{1t},\, \ldots,\, u_{2t}]'$ of residuals such that:

$$u_t | \mathcal{F}_{t-1} \sim \mathcal{N}\left(0,\, \Sigma_t\right), \tag{1.35}$$

with

$$\Sigma_t = D_t^{\frac{1}{2}} R_t D_t^{\frac{1}{2}} \tag{1.36}$$

following a law such as

$$\Sigma_t = \mathcal{V}_{t-1}\left(u_t\right), \tag{1.37}$$

so that

$$D_t = diag\left\{\Sigma_t\right\}, \tag{1.38}$$

is a $K \times K$ diagonal matrix of conditional variances.

In the DCC framework, the dynamics of variances and correlations are specified separately. A two-step procedure (three steps when covariance targeting is employed) is then used in the estimation of the parameters. In Engle and Sheppard [15] and Engle [12] it is in fact shown that the parameters for the dynamics of variances and those for the dynamics of correlations can be consistently estimated via two distinct ML optimizations.

The first optimization recovers the parameters for the variances process. Once these parameters are obtained, a second ML optimization, conditional on the former, delivers the parameters for the dynamics of the correlations.

Still following Engle [13], conditional variances are modeled in K GARCH(1,1) equations:

$$\sigma_{ii,\,t} = \omega_i + \alpha_i y_{i,\,t-1}^2 + \beta_i \sigma_{ii,\,t-1}. \tag{1.39}$$

Note that such a modeling strategy does not accommodate for the leverage effect.

Finally, correlations are modeled by first DE-GARCHING the raw residuals. This procedure is performed by simply dividing the raw residuals by the conditional variances:

$$\varepsilon_t = D_t^{-\frac{1}{2}} y_t. \tag{1.40}$$

Then, it is possible to define a process Q_t, such that its dynamics follow, in terms of the standardized residuals, a diagonal scalar GARCH(1,1):

$$Q_t = \Omega + \alpha \varepsilon_{t-1} \varepsilon'_{t-1} + \beta Q_{t-1}. \tag{1.41}$$

To ensure that correlations fall between the interval $(-1,\ 1)$ the following normalization is carried out:

$$R_t = diag\left\{Q_t\right\}^{-\frac{1}{2}} Q_t diag\left\{Q_t\right\}^{-\frac{1}{2}}. \tag{1.42}$$

Estimation of the GARCH(1, 1) Scalar Diagonal DCC Model

Given a zero mean conditionally heteroscedastic multivariate normal process and a sample with width T, the log likelihood for the GARCH(1, 1) Scalar Diagonal DCC Model can be written as:

$$\begin{aligned}
\mathcal{L} &= -\frac{1}{2} \sum_{t=1}^{T} \left(K \log\left(2\pi\right) + \log\left|\Sigma_t\right| + u'_t \Sigma_t^{-1} u_t \right) \\
&= -\frac{1}{2} \sum_{t=1}^{T} \left(K \log\left(2\pi\right) + \log\left|D_t^{\frac{1}{2}} R_t D_t^{\frac{1}{2}}\right| + u'_t D_t^{-\frac{1}{2}} R_t^{-1} D_t^{-\frac{1}{2}} u_t \right) \\
&= -\frac{1}{2} \sum_{t=1}^{T} \left(K \log\left(2\pi\right) + \log\left|D_t\right| + \log\left|R_t\right| + \varepsilon'_t R_t^{-1} \varepsilon_t \right) \\
&= -\frac{1}{2} \sum_{t=1}^{T} \left(K \log\left(2\pi\right) + \log\left|D_t\right| + u'_t D_t^{-1} u_t + \varepsilon'_t \varepsilon_t + \log\left|R_t\right| + \varepsilon'_t R_t^{-1} \varepsilon_t \right) \\
&= \mathcal{L}_1 + \mathcal{L}_2 + \varepsilon'_t \varepsilon_t,
\end{aligned} \tag{1.43}$$

where

$$
\begin{aligned}
\mathcal{L}_1 &= -\frac{1}{2}\sum_{t=1}^{T}\left(K\log(2\pi)+\log|D_t|+u_t'D_t^{-1}u_t\right)\\
&= -\frac{1}{2}\sum_{t=1}^{T}\left(K\log(2\pi)+\sum_{i=1}^{K}\left(\log(\sigma_{ii,\,t})+\frac{u_{it}^2}{\sigma_{ii,\,t}}\right)\right)\\
&= -\frac{1}{2}\sum_{i=1}^{K}\left(K\log(2\pi)+\sum_{t=1}^{T}\left(\log(\sigma_{ii,\,t})+\frac{u_{it}^2}{\sigma_{ii,\,t}}\right)\right), \qquad (1.44)
\end{aligned}
$$

is the sum of the univariate GARCH likelihoods, which can be maximized individually and

$$
\mathcal{L}_2 = -\frac{1}{2}\sum_{t=1}^{T}\left(\log|R_t|+\varepsilon_t'R_t^{-1}\varepsilon_t\right), \qquad (1.45)
$$

maximized by imposing

$$
Q_t = \Omega + \alpha\varepsilon_{t-1}\varepsilon_{t-1}' + \beta Q_{t-1} \qquad (1.46)
$$

and

$$
R_t = diag\left\{Q_t\right\}^{-\frac{1}{2}} Q_t diag\left\{Q_t\right\}^{-\frac{1}{2}}. \qquad (1.47)
$$

As Engle [13] suggests, the term in the sum of squared epsilons can be ignored as it does not depend on the parameters being optimised.

1.3.2 Forecasting Variance, Correlation and Variance Covariance Matrices with a GARCH(1, 1) Scalar Diagonal DCC Model

As variances are GARCH(1, 1) the one step ahead forecast of variances, conditional to information in t is

$$
\begin{aligned}
\sigma_{ii,\,t}\left(1\right) &= \left(1-\alpha-\beta\right)\sigma_{ii}+\alpha u_{ii,\,t}^{2}+\beta\sigma_{ii,\,t} \\
&= \sigma_{ii}+\alpha\left(u_{ii,\,t}^{2}-\sigma_{ii}\right)+\beta\left(\sigma_{ii,\,t}-\sigma_{ii}\right),
\end{aligned}
\tag{1.48}
$$

where σ_{ii} is an unconditional variance and it can be checked that

$$
\sigma_{ii,\,t}\left(H-h\right)=\sigma_{ii}+\left(\alpha+\beta\right)^{H-1-h}\left(\sigma_{ii,\,t}\left(1\right)-\sigma_{ii}\right).
\tag{1.49}
$$

1.48 can be vectorized, so that variance matrix process follows

$$
\begin{aligned}
D_t\left(1\right) &= \left(1-\alpha-\beta\right)D+\alpha u_t u_t'+\beta D_t \\
&= D+\alpha\left(u_t^2-D\right)+\beta\left(D_t-D\right),
\end{aligned}
\tag{1.50}
$$

with D being an unconditional variance matrix and

$$
D_t\left(H-1-h\right)=D+\left(\alpha+\beta\right)^{H-1-h}\left(D_t\left(1\right)-D\right).
\tag{1.51}
$$

Engle [13] shows that for $h>1$ it is not possible to produce an exact analytical expression for the forecasts of the correlation matrix. In order to overcome this obstacle, it is proposed an assumption suggested by Engle and Sheppard [15], that is

$$
R_t\left(h\right)\approx Q_t\left(h\right),
\tag{1.52}
$$

the approximation being as precise as the diagonal elements of Q are all close to one. Then,

$$
R_t\left(h\right)=\bar{R}+\left(\alpha+\beta\right)\left(R_t\left(h-1\right)-\bar{R}\right).
\tag{1.53}
$$

Finally the forecast of the conditional variance covariance matrix can be estimated as:

$$\Sigma_t(h) \approx D_t(h)^{\frac{1}{2}} R_t(h) D_t(h)^{\frac{1}{2}}. \tag{1.54}$$

1.4 Assessment of Heteroscedastic Effects

1.4.1 OVD Based Measures of Connectedness

Consider again the system

$$y_t = \nu + \sum_{l=1}^{p} A_l y_{t-l} + u_t. \tag{1.55}$$

If u_t is still serially uncorrelated, but displays conditional heteroscedasticity the OMCs must be adapted in the following way. For every sorting of the system variables, at each point in time, the Cholesky lower triangular decomposition of the conditional variance covariance matrix is:

$$\Sigma_t = P_t P_t'. \tag{1.56}$$

It follows that, if we define

$$w_t = P_t^{-1} u_t, \tag{1.57}$$

w_t is still an error term with mean zero and

$$E(w_t w_t') = \Sigma_w = E\left(P_t^{-1} u_t u_t' \left(P_t^{-1}\right)'\right) = P_t^{-1} \Sigma_t \left(P_t^{-1}\right)' = I_K. \tag{1.58}$$

The MA process must then be rewritten as

$$y_t = \dot{\Theta}(L)\, w_t, \tag{1.59}$$

where

$$\dot{\Theta}(L) = \left(\dot{\Theta}_t(0) + \dot{\Theta}_t(-1)\, L + \dot{\Theta}_t(-2)\, L^2 + \ldots\right) = \left(P_t + \Phi_1 P_{t-1} + \Phi_2 P_{t-2} + \ldots\right) \tag{1.60}$$

and the dot notation means that it accounts for heteroscedasticity.

The H-steps ahead forecast will be equal to:

$$\begin{aligned} \dot{y}_t(H) &= \mu + \sum_{h=H}^{\infty} \Phi_h u_{t+H-h} \\ &= \mu + \sum_{h=H}^{\infty} \Phi_h P_t(H-h)\left(P_t(H-h)\right)^{-1} u_{t+H-h} \\ &= \mu + \sum_{h=H}^{\infty} \dot{\Theta}_t(h)\, w_{t+H-h}, \end{aligned} \tag{1.61}$$

with forecast error

$$\begin{aligned} \dot{\xi}_t(H) &= y_{t+h} - y_t(H) \\ &= \sum_{h=0}^{\infty} \dot{\Theta}_t(h)\, w_{t+H-h} - \sum_{h=H}^{\infty} \dot{\Theta}_t(h)\, w_{t+H-h} \\ &= \sum_{h=0}^{H-1} \dot{\Theta}_t(h)\, w_{t+H-h}. \end{aligned} \tag{1.62}$$

As the elements of w are still not correlated,

$$\text{MSE}\left[\dot{y}_t(H)\right] = E_t\left[\dot{\xi}_t(H)\, \dot{\xi}_t'(H)\right]$$

$$= \sum_{h=0}^{H-1} \dot{\Theta}_t(h) \left(\dot{\Theta}_t(h) \right)'. \tag{1.63}$$

Then, explicitly accounting for heteroscedasticity, the original method of Diebold and Yilmaz [8] can be adjusted as:

$$\begin{aligned} \dot{d}_{t,\,ij}^{(OVD)}(H) &= \frac{\sum_{h=0}^{H-1} \left(e_i' \dot{\Theta}_t(h) e_j \right)^2}{\sum_{h=0}^{H-1} e_i' \dot{\Theta}_t(h) \left(\dot{\Theta}_t(h) \right)' e_i} \\ &= \frac{\sum_{h=0}^{H-1} \dot{\theta}_{t,\,ij}^2(h)}{\sum_{h=0}^{H-1} \sum_{j=1}^{K} \dot{\theta}_{t,\,ij}^2(h)} \end{aligned} \tag{1.64}$$

and the heteroscedasticity adjusted OMTC can be computed as

$$\begin{aligned} \dot{S}_t^{(OVD)}(H) &= \frac{\sum_{i,j=1,\,i\neq j}^{K} \dot{d}_{t,\,ij}^{(OVD)}(H)}{\sum_{i,j=1}^{K} \dot{d}_{ij}^{(OVD)}(H)} \\ &= \frac{1}{K} \sum_{i,j=1,\,i\neq j}^{K} \dot{d}_{t,\,ij}^{(OVD)}(H). \end{aligned} \tag{1.65}$$

1.4.2 GVD Based Measures of Connectedness

Given the usual system

$$y_t = \nu + \sum_{l=1}^{p} A_l y_{t-l} + u_t, \tag{1.66}$$

conditional on the information in t and the shocks to the i-th equation up to time $t+H$, that is, $u_{it},\, u_{i,\,t+1},\, \ldots,\, u_{i,\,t+H}$, the H-steps ahead forecast will be:

$$y_t^{(i)}(H) = \mu + \sum_{h=H}^{\infty} \Phi_h u_{t+H-h} + \sum_{h=0}^{H-1} E_t\left(u_{t+H-h} \middle| u_{i,\,t+H-h} \right). \tag{1.67}$$

Assuming that $u_t \sim N(0,\, \Sigma_t)$, if u_t is conditionally heteroscedastic,

$$E_t\left(u_{t+H-h}\right|u_{i,\,t+H-h}\right) = \left((\sigma_{ii,\,t}\,(H-h))^{-1}\,\Sigma_t\,(H-h)\,e_i\right)u_{i,\,t+H-h}, \tag{1.68}$$

where $\sigma_{ii,\,t}\,(t+H-h)$ and $\Sigma_t\,(t+H-h)$ are the $(H-h)$-steps ahead forecasts of the standard deviation of $u_{i,\,t}$ and the variance covariance matrix of u_t, both conditional, for tractability, only on the information available at time t, making 1.68 an approximate relationship.

Then, e_i still being the i-th column of an identity matrix of order K,

$$\begin{aligned}\dot{y}_t^{(i)}\,(H) = \mu &+ \sum_{h=H}^{\infty} \Phi_h u_{t+H-h} \\ &+ \sum_{h=0}^{H-1} \Phi_h\left((\sigma_{ii,\,t}\,(H-h))^{-1}\,\Sigma_t\,(H-h)\,e_i\right)u_{i,\,t+H-h},\end{aligned} \tag{1.69}$$

Along these lines, it follows that the H-steps ahead forecast errors will be

$$\begin{aligned}\dot{\xi}_t^{(i)}\,(H) &= y_{t+h} - y_t^{(i)}\,(H) \\ &= \sum_{h=0}^{H-1} \Phi_h u_{t+H-h} \\ &- \sum_{h=0}^{H-1} \Phi_h\left((\sigma_{ii,\,t}\,(H-h))^{-1}\,\Sigma_t\,(H-h)\,e_i\right)u_{i,\,t+H-h} \\ &= \sum_{h=0}^{H-1} \Phi_h\left[u_{t+H-h} - \left((\sigma_{ii,\,t}\,(H-h))^{-1}\,\Sigma_t\,(H-h)\,e_i\right)u_{i,\,t+H-h}\right].\end{aligned} \tag{1.70}$$

In Appendix 1.C it is shown that

$$MSE\left[\dot{y}_t^{(i)}\,(H)\right] = E_t\left[\dot{\xi}_t^{(i)}\,(H)\,\dot{\xi}_t^{(i)\prime}\,(H)\right]$$

$$
\begin{aligned}
&= \sum_{h=0}^{H-1} \Phi_h \Sigma_t \left(t + H - h\right) \Phi_h' \\
&- \left(\sum_{h=0}^{H-1} \left(\sigma_{ii,t}\left(H-h\right)\right)^{-1} \Phi_h \Sigma_t \left(H-h\right) e_i e_i' \Sigma_t \left(H-h\right) \Phi_h' \right) \\
&= \sum_{h=0}^{H-1} \Phi_h \Sigma_t \left(H - h\right) \Phi_h' \\
&- \sum_{h=0}^{H-1} \left(\sigma_{ii,t}\left(H-h\right)\right)^{-1} \left(\Phi_h \Sigma_t \left(H-h\right) e_i\right)^2 .
\end{aligned}
\tag{1.71}
$$

Finally, define

$$
\begin{aligned}
\dot{\Delta}_{t,i}\left(H\right) &= MSE\left[\dot{\xi}_t\left(H\right)\right] - MSE\left[\dot{\xi}_t^{(i)}\left(H\right)\right] \\
&= \left(\sigma_{ii,t}\left(H-h\right)\right)^{-1} \left(\sum_{h=0}^{H-1} \Phi_n \Sigma_t \left(H-h\right) e_i \right)^2 ,
\end{aligned}
\tag{1.72}
$$

the extension of the GVD for heteroscedasticity returns

$$
\dot{d}_{t,ij}^{(GVD)}\left(H\right) = \frac{\sum_{h=0}^{H-1} \sigma_{ii,t}^{-1}\left(H-h\right)\left(\Phi_h \Sigma_t \left(H-h\right) e_i\right)^2}{\sum_{h=0}^{H} e_j' \Phi_h \Sigma_t \left(H-h\right) \Phi_n' e_j}, \tag{1.73}
$$

where the superscript signifies it is the generalized version of the measure, which is conditional in t, from i to j, and with a forecast horizon equal to H periods ahead.

Since the GVD includes effects due to covariances (unless the variance covariance matrix of u is diagonal), forecast error variance contributions do not sum to one, as the sum is not the MSE of the i-th equation. To standardize the measure, one can adopt a normalization that depends on the focus of the analysis. For instance, if one wants wants to obtain a percentage of the variance spread from each variable through the rest of the system, the following normalization can be used:

$$
\tilde{\dot{d}}_{t,ij}^{(GVD)}\left(H\right) = \frac{\dot{d}_{t,ij}^{(GVD)}\left(H\right)}{\sum_{i=1}^{K} \dot{d}_{t,ij}^{(GVD)}\left(H\right)}. \tag{1.74}
$$

By construction then, $\sum_{i=1}^{K} \tilde{\dot{d}}_{t,\,ij}^{(GVD)}\,(H) = 1$ and $\sum_{i,\,j=1}^{K} \tilde{\dot{d}}_{t,\,ij}^{(GVD)}\,(H) = N$.

To obtain a percentage index of the variance received from other variables, the normalizations to use will be:

$$\tilde{\dot{d}}_{t,\,ij}^{(GVD)}\,(H) = \frac{\dot{d}_{t,\,ij}^{(GVD)}\,(H)}{\sum_{j=1}^{K} \dot{d}_{t,\,ij}^{(GVD)}\,(H)}. \tag{1.75}$$

The Generalized Measure of Total Connectedness (GMTC) can then be computed as

$$\begin{aligned}\dot{S}_{t}^{(GVD)}\,(H) &= \frac{\sum_{i,\,j=1,\,i\neq j}^{K} \tilde{\dot{d}}_{t,\,ij}^{(GVD)}\,(H)}{\sum_{i,\,j=1}^{K} \tilde{\dot{d}}_{t,\,ij}^{(GVD)}\,(H)} \\ &= \frac{1}{K} \sum_{i,\,j=1,\,i\neq j}^{K} \tilde{\dot{d}}_{t,\,ij}^{(GVD)}\,(H)\,, \end{aligned} \tag{1.76}$$

which provides the same number independently from the normalization scheme chosen.

In the Appendix 1.D, the reader could find a formulation optimized for computer programming.

1.5 Empirical Illustration

World markets are connected through a complex web of trade and financial relationships. Understanding this connectedness would require an extensive use of macroeconomic models and knowledge of a vast number of balance sheets. One way to simplify the problem would be to exploit the information contained in a dataset of international indices and apply the Diebold and Yilmaz Framework to analyze the connectedness arising from it. An important feature of the following analysis is that it remains agnostic about how the connectedness arises and the sole scope is to measure it and its evolution.

As Diebold and Yilmaz [9] point out, the study of volatility connectedness is particularly useful, because it allows to assess the connectedness of uncertainty ("fear") and because volatility is particularly crises-sensitive. As the world is recovering from an exceptionally severe economic and financial crisis, this seemed an interesting case to be studied. Moreover, volatilities time series could be characterized by conditional heteroscedasticity; to enable the assessment of when and how this is relevant in the Diebold-Yilmaz framework, this paper extends it to explicitly allow for heteroscedasticity and embeds this case both at a static and at a dynamic level.

Volatilities are latent and must be estimated. This paper will employ the Heber, Gerd, Asger Lunde, Neil Shephard and Kevin Sheppard (2009) "Oxford-Man Institute's Realized Library", Oxford-Man Institute, University of Oxford in its 2.0 version. In summary, among other, the library contains time series of daily realized variances of a number of international indices. Of the the entire dataset above specified, we will use the series listed in Table 1.1.

[Table 1.1 about here.]

The dataset starts at 03:01:2000 and ends at 01:05:2013 (ISO 8601 Date Format). Table 1.1 also reports the list of the variables present in the original dataset.

The original dataset is expressed at a daily frequency. In order to get rid of some microstructure effects, we aggregate the data at a weekly level, having the week starting on Wednesday. As some data points will be missing, the resulting numbers will be averages of a variable number of entries, which means substituting the missing value with the week average.

The actual time series used in the analysis will be the series of log transformations of volatilities, as shown by Andersen et al. [1] to be suitable to a normal approximation, which, in turn, is a condition required by the generalized variance decompositions (Koop et al. [20], Pesaran and Shin [23], Pesaran and Pesaran [22]).

As volatilities tend to display strong serial correlation, the subset of the transformed time series will be modeled as a VAR(p) and the measures of connectedness will be computed from both the forecast errors and the error terms expressed by the model. The forecast horizon is set at an arbitrary five weeks. The order of the vector autoregression is chosen by the HIC for the full sample, setting the maximum autoregressive order required by the algorithm at five periods.

In order to capture the nonlinearity in the dataset, following the Diebold and Yilmaz literature, a number of rolling windows estimations are performed, the window width being somewhat arbitrarily set at one hundred observations, which will set the initial rolling windows measurements in the window starting from week 2000:01 and ending at 2001:47. Setting the window width entails a tradeoff between the probability of observing heteroscedastic effects (volatility tends to cluster) and the number of rolling windows measurements. The decision taken in this paper is to set a window wide enough to contain heteroscedastic episodes with acceptable probability. Finally, although this could be a debatable choice, the order of the vector autoregression will be maintained during rolling windows estimates, because it will lower computation time by a substantial amount and would allow weaker signals of autoregressive effects to be captured.

The result of the latter procedure will return a series of time varying measures of connectedness. To assess the impact that heteroscedasticity has in this variation, for each subsample, the specification is extended to explicitly model it via the Diagonal Scalar DCC Model. To test for heteroscedasticity, the Engle [14] test has been used for each time series. Such test fails to reject the hypothesis of heteroscedasticity for all series, at least given the full sample. Every time the optimization fails for any subsample, we consider the latter as displaying homoscedasticity.

Finally, a smoothed, full sample, version of heteroscedastic connectedness will be computed, in which only the variance covariance matrix will be allowed to vary according to a Diagonal Scalar DCC Model. The measure is then computed using the realized values of the covariance matrices H periods ahead of each sample point.

The following analysis loosely follows the analytic structure of Diebold and Yilmaz [9]; the main object will be to sketch whether, when and how the connectedness measures will be affected by the explicit modeling of heteroscedasticity.

1.5.1 Static (Full-Sample, Unconditional) Analysis

Here, connectedness measures will be computed using the information included in the whole sample. The data is modeled with a VAR(p) and both homoscedastic and heteroscedastic cases are studied, both in the orthogonal and in the generalized cases. In this analysis, connectedness will be computed only at the end of the whole sample. The heteroscedastic case differentiates from the homoscedastic one by allowing the covariances inside the connectedness formulae to be variant and, specifically, be the forecasts outlined in section 1.3.

Tables 1.2 and 1.3 report connectedness matrices for orthogonal homoscedastic connectedness and orthogonal heteroscedastic connectedness.

[Table 1.2 about here.]

[Table 1.3 about here.]

Tables 1.4 and 1.5 report connectedness matrices for generalized homoscedastic connectedness and generalized heteroscedastic connectedness.

[Table 1.4 about here.]

[Table 1.5 about here.]

Tables 1.6 and 1.7 summarize changes occurring from the homoscedastic to the heteroscedastic case for both the orthogonal and the generalized measures of connectedness, respectively.

[Table 1.6 about here.]

[Table 1.7 about here.]

Considering the latter tables, it is immediately possible to observe that heteroscedasticity could be indeed an important feature when analyzing connectedness measures. Even though the change in total connectedness is not much higher (in absolute value) than four percentage points, heteroscedasticity plays a bigger role, when considering changes in “from“ and “to” measures. In fact, we can observe striking changes of more than 40% (in absolute value), when considering the change in the orthogonal to measures and changes of up to 30% (in absolute value) in the generalized from measures.

From a robustness point of view, the heteroscedastic case maintains the feature by which the generalized version of total connectedness acts as an upper bound for the orthogonal case. This effect is due to the fact that the generalized case, by treating each variable in the system as the first one in a WO (Kim [19]), includes, in the computation of the measure, covariances that would be absent in the orthogonal case. Such robustness is not preserved in disaggregate measures of connectedness both in the homoscedastic and the heteroscedastic cases.

1.5.2 Dynamic Analysis

Diebold and Yilmaz [9] already pointed out that measuring and frequent monitoring of connectedness measures can help understanding the evolution of financial facts, in particular crises. In fact, total connectedness can be seen as a measure of systemic risk, as it is a measure of variance spilled over from and to the system equations. This section will describe how the explicit modeling of heteroscedastic effects could influence these measures in a dynamic setting.

In this analysis of the dynamics originated by the nonlinearities in the data, we use two routes. The first route, consistent with the previous body of work on connectedness, consists in performing a batch of rolling windows estimations, in which both the parameters of the VAR and the Diagonal Scalar DCC Model are allowed to vary and are estimated with the sample included in each window. For every window of estimation, a value for the connectedness matrix is estimated. Time varying variances and correlations will affect the forecasts needed for the computation of the measures.

In the second route, the parameters of both the VAR and the Diagonal Scalar DCC Model are forced to be constant and estimated using the whole sample. With the estimated model, connectedness measures are then computed using the in sample conditional dynamic covariance matrices expressed by the Diagonal Scalar DCC Model in the formulae for connectedness, outputting, thus, a smoothed version of the DY measures.

Rolling Windows Analysis

Figures 1.1 and 1.2 report plots of rolling windows total connectedness measures, both homo- and heteroscedastic.

[Figure 1.1 about here.]

[Figure 1.2 about here.]

Figure 1.1 shows orthogonal measures, Figure 1.2 generalized measures of total connectedness.

The first thing to note is that in both pictures homoscedastic and heteroscedastic measures of total connectedness tend to follow a very similar path. This could be due to the fact that, in small samples, in sample heteroscedastic effects could assume a second order impact in the manipulation of the information embedded in the data. This notwithstanding, a discrepancy from the homoscedastic benchmark signifies episodes of abnormal activity in the covariance domain.

Figures 1.3 and 1.4 show time series of deviations from the homoscedastic measures taken by the heteroscedastic version of the total connectedness measures, orthogonal and generalized, respectively.

[Figure 1.3 about here.]

[Figure 1.4 about here.]

It is interesting to see that these deviations appear in the same periods for both the orthogonal and the generalized version of the measures, magnified in the generalized case, as it provides an upper bound to the orthogonal total connectedness.

These spikes seem to occur during episodes of particular crises in the financial markets, possibly identifiable around the Worldcom scandal of 2002, the beginning of the credit crisis in 2006, the outburst of the Greek crisis at the end of 2009 and a peak of the Euro crisis at the end of 2010.

Heteroscedastic Full Sample Analysis

In this case, parameters for both the VAR and the Diagonal Scalar DCC Model are kept constant and estimated with the full sample, while dynamics in connectedness are generated solely by the smoothed conditional dynamic covariances. Figure 51.5 shows the plots for both the orthogonal and generalized total connectedness, the latter maintaining the role of upper bound.

[Figure 1.5 about here.]

It is immediately evident the difference between this picture and those obtained in the rolling windows case, as the only source of variability is the dynamic covariance matrix, while the other parameters have been averaged out in the full sample.

The second interesting feature of these plots is that they appear to display a two regimes pattern: a pre financial crisis regime characterised by low connectedness and crisis regime characterised by high connectedness, with the structural break seamingly happening in the middle of 2006 and a possible, interesting and optimistic, second break happening in the middle of 2013. A regime switching model could be applied to analyze whether this is true.

A Comment on Rolling Windows and Heteroscedastic Full Sample Analyses

In this paper, both rolling windows and heteroscedastic full sample analyses have been used to extract information on connectedness from the data. It is important to note that each of them provides specific information not available to the other technique.

Rolling windows allow all the parameters of the model to vary, which in turn makes the measure of total connectedness more erratic and allows certain patterns to emerge. For instance, it is possible to see in Figures 1.1 and 1.2 two inverse parabola shapes characterizing two cycles on top of the two regimes described by Figure 1.5. Heteroscedasticity introduced in the rolling windows analysis allows even finer turbulence in the measure to be spotted.

On the other hand, the heteroscedastic full sample analysis, averaging the dynamics of all the parameters, with the exception of the covariance matrix, allows the emergence of macro patterns like the two regimes displayed by Figure 1.5.

The different time series generated by these two techniques could then be suited to further treatment and be the basis of more econometric analysis. For instance, the time series in Figure 1.5 could be used to estimate a two regimes switching model that can be applied to infer structural breaks, while the more erratic time series of Figures 1.1 and 1.2 could be used to infer the existence of cyclicality in total connectedness.

1.6 Conclusions

The recent financial crisis has underlined the importance of interdependence in the development of financial turbulence. One way to characterize such interdependence is the connectedness framework introduced by Diebold and Yilmaz [8] and extended by Diebold and Yilmaz [9]. In this paper, we extended the connectedness framework explicitily allowing for conditional heteroscedasticity and we applied the model to analyze connectedness in the volatility of international equity indexes.

Both static and dynamic analyses have been performed and in all cases has emerged a positive value in extending the framework to explicitly allow for the modeling of heteroscedasticity. This value substantiates in a finer characterization of financial events such as the Worldcom scandal, the beginning of the credit crisis and key episodes of the crisis of the Eurozone. Moreover, moving outside of the rolling windows analysis and applying a novel heteroscedastic full sample analysis, it was possible to allow for the emergence of a two regimes patterns characterizing, consistently with recent economic history, a regime with low connectedness and a regime with high connectedness, possibly ending in 2013. At the time of writing (December 2013), the Fed recently announced a plan to slow bond purchases.

Bibliography

[1] Andersen, T. G., Bollerslev, T., Diebold, F. X., and Labys, P. (2003). Modeling and forecasting realized volatility. *Econometrica*, 71(2):579–625.

[2] Antonakakis, N. (2012). Exchange return co-movements and volatility spillovers before and after the introduction of euro. *Journal of International Financial Markets, Institutions and Money*, 22(5):1091–1109.

[3] Antonakakis, N., Floros, C., and Kizys, R. (2015). Dynamic spillover effects in futures markets: Uk and us evidence. *International Review of Financial Analysis*.

[4] Bollerslev, T., Engle, R. F., and Wooldridge, J. M. (1988). A capital asset pricing model with time-varying covariances. *Journal of political Economy*, 96(1):116–131.

[5] Bostanci, G. and Yilmaz, K. (2015). How connected is the global sovereign credit risk network?

[6] Demirer, M., Diebold, F. X., Liu, L., and Yılmaz, K. (2017). Estimating global bank network connectedness. Technical report, National Bureau of Economic Research.

[7] Diebold, F. X., Liu, L., and Yilmaz, K. (2017). Commodity connectedness.

[8] Diebold, F. X. and Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119(534):158–171.

[9] Diebold, F. X. and Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1):57–66.

[10] Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134.

[11] Diebold, F. X. and Yilmaz, K. (2015). *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press, USA.

[12] Engle, R. (2002). Dynamic conditional correlation: A simple class of multivariate generalized autoregressive conditional heteroskedasticity models. *Journal of Business & Economic Statistics*, 20(3):339–350.

[13] Engle, R. (2009). *Anticipating correlations: a new paradigm for risk management*. Princeton University Press.

[14] Engle, R. F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of united kingdom inflation. *Econometrica: Journal of the Econometric Society*, pages 987–1007.

[15] Engle, R. F. and Sheppard, K. (2001). Theoretical and empirical properties of dynamic conditional correlation multivariate garch. Technical report, National Bureau of Economic Research.

[16] Forbes, K. J. and Rigobon, R. (2002). No contagion, only interdependence: measuring stock market comovements. *The journal of Finance*, 57(5):2223–2261.

[17] Gaspar, R. (2012). Comment on better to give than to receive by francis x. diebold and kamil yilmaz. 28.

[18] Hamilton, J. D. (1994). *Time series analysis*. Princeton university press Princeton.

[19] Kim, H. (2009). Generalized impulse response analysis: general or extreme?

[20] Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, 74(1):119–147.

[21] Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

[22] Pesaran, B. and Pesaran, M. H. (2010). *Time series econometrics using Microfit 5.0: A user's manual*. Oxford University Press, Inc.

[23] Pesaran, H. H. and Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics letters*, 58(1):17–29.

[24] Wold, H. O. (1960). A generalization of causal chain models (part iii of a triptych on causal chain systems). *Econometrica: Journal of the Econometric Society*, pages 443–463.

Appendix 1.A The Cholesky Algorithm

This appendix discusses the Cholesky decomposition. The exposition will be brief and compressed; for further details refer to Hamilton [18].

The Cholesky decomposition is based on the triangular decomposition of a symmetric positive definite square matrix. The variance covariance matrix is a $K \times K$ matrix defined as such. A way to state the triangular decomposition is the following:

$$\Sigma = T\Sigma^{(K)}T', \tag{A.1}$$

where the term in the parenthesis of the suffix in Σ represents the iteration of the algorithm and

$$T = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ t_{21} & 1 & 0 & \dots & 0 \\ t_{31} & t_{32} & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ t_{K1} & t_{K2} & t_{K4} & \dots & 1 \end{bmatrix}, \tag{A.2}$$

$$\Sigma^{(K)} = \begin{bmatrix} \sigma_{11}^{(K)} & 0 & 0 & \cdots & 0 \\ 0 & \sigma_{22}^{(K)} & 0 & \dots & 0 \\ 0 & 0 & \sigma_{33}^{(K)} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & \sigma_{KK}^{(K)} \end{bmatrix}. \tag{A.3}$$

The procedure is indeed iterative and can be summarized with the following recursion:

$$\Sigma^{(j)} = T^{(j-1)}\Sigma^{(j-1)}T^{(j-1)},\ \forall j = 1,\, 2,\, \dots,\, K, \tag{A.4}$$

$$
\begin{aligned}
T &= \prod_{i=1}^{K-1} \left(T^{(i)}\right)^{-1} \\
&= \left[\left(T^{(1)}\right)^{-1}_{\bullet,\,1} \left(T^{(2)}\right)^{-1}_{\bullet,\,2} \ldots \left(T^{(K-1)}\right)^{-1}_{\bullet,\,K-1} e_K \right],
\end{aligned}
\tag{A.5}
$$

where $\left(T^{(i)}\right)^{-1}_{\bullet,\,j}$ is the j-th column of $\left(T^{(i)}\right)^{-1}$ and

$$
\Sigma^{(1)} = \Sigma. \tag{A.6}
$$

The iteration starts at $j = 2$ and the transformation matrix $T^{(j)}$ is such that it has ones on the diagonal and zeros in all the other positions, exept in the $(j-1) - th$ column at positions $(i,\, j-1 \,|\, i > j-1)$. The quantities different from zero or one can be expressed as:

$$
t^{(j-1)}_{i,\,j-1} \,|\, i > j-1 = -\sigma^{(j-1)}_{i,\,j-1} \left[\sigma^{(j-1)}_{j-1,\,j-1}\right]^{-1}. \tag{A.7}
$$

The conditioned matrix $\Sigma^{(j)}$, for $j \geq 2$ is such that

$$
\sigma^{(j)}_{ik} \mid (i \leq j-1 \,\vee\, k \leq j-1) = 0, \tag{A.8}
$$

$$
\sigma^{(j)}_{ii} \,|\, i \leq j = \sigma^{(i)}_{ii}, \tag{A.9}
$$

$$
\sigma^{(j)}_{ik} \mid (i \geq j \,\wedge\, k \geq j) = \sigma^{(j-1)}_{ik} - \sigma^{(j-1)}_{i,\,j-1} \left[\sigma^{(j-1)}_{j-1,\,j-1}\right]^{-1} \sigma^{(j-1)}_{j-1,\,k}. \tag{A.10}
$$

For example, if $j = 2$, such that $\Sigma^{(2)} = T^{(1)}\Sigma^{(1)} \left[T^{(1)}\right]'$,

$$T^{(1)} = \begin{bmatrix} 1 & 0 & 0 & \dots & 0 \\ -\sigma_{21}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} & 1 & 0 & \dots & 0 \\ -\sigma_{31}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\sigma_{K1}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} & 0 & 0 & \dots & 1 \end{bmatrix} \tag{A.11}$$

and

$$\Sigma^{(2)} = \begin{bmatrix} \sigma_{11}^{(1)} & 0 & 0 & \dots & 0 \\ 0 & \sigma_{22}^{(1)} - \sigma_{21}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} \sigma_{12}^{(1)} & \sigma_{23}^{(1)} - \sigma_{21}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} \sigma_{13}^{(1)} & \dots & \sigma_{2K}^{(1)} - \sigma_{21}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} \sigma_{1K}^{(1)} \\ 0 & \sigma_{32}^{(1)} - \sigma_{31}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} \sigma_{12}^{(1)} & \sigma_{33}^{(1)} - \sigma_{31}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} \sigma_{13}^{(1)} & \dots & \sigma_{2K}^{(1)} - \sigma_{31}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} \sigma_{1K}^{(1)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \sigma_{K2}^{(1)} - \sigma_{K1}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} \sigma_{12}^{(1)} & \sigma_{K3}^{(1)} - \sigma_{K1}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} \sigma_{13}^{(1)} & \dots & \sigma_{KK}^{(1)} - \sigma_{K1}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} \sigma_{1K}^{(1)} \end{bmatrix}. \tag{A.12}$$

As mentioned before, the Cholesky decomposition is based on the triangular factorization of a symmetric positive definite square matrix. To obtain it, first define $\left[\Sigma^{(K)}\right]^{\frac{1}{2}}$ as the $(K \times K)$ matrix whose diagonal entries are the square roots of the diagonal elements of $\Sigma^{(K)}$. Then,

$$\begin{aligned} \Sigma &= T\Sigma^{(K)}T' \\ &= T\left[\Sigma^{(K)}\right]^{\frac{1}{2}}\left[\Sigma^{(K)}\right]^{\frac{1}{2}}T' \\ &= \left\{T\left[\Sigma^{(K)}\right]^{\frac{1}{2}}\right\}\left\{T\left[\Sigma^{(K)}\right]^{\frac{1}{2}}\right\}' \\ &= PP'. \end{aligned} \tag{A.13}$$

and P is the lower Cholesky triangular decomposition having the following form:

$$P = \begin{bmatrix} \sqrt{\sigma_{11}^{(1)}} & 0 & \dots & 0 \\ \sigma_{21}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} \sqrt{\sigma_{11}^{(1)}} & \sqrt{\sigma_{22}^{(2)}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K1}^{(1)} \left[\sigma_{11}^{(1)}\right]^{-1} \sqrt{\sigma_{11}^{(1)}} & \sigma_{K2}^{(2)} \left[\sigma_{22}^{(2)}\right]^{-1} \sqrt{\sigma_{22}^{(2)}} & \dots & \sqrt{\sigma_{KK}^{(K)}} \end{bmatrix}, \tag{A.14}$$

so that

$$p_{ik} \,|\, i < k = 0, \tag{A.15}$$

$$p_{ii} = \sqrt{\sigma_{ii}^{(i)}}, \tag{A.16}$$

$$\begin{aligned} p_{ik} \,|\, i > k &= \sigma_{ik}^{(k)} \left[\sigma_{kk}^{(k)}\right]^{-1} \sqrt{\sigma_{kk}^{(k)}} \\ &= \frac{\sigma_{ik}^{(k)}}{\sqrt{\sigma_{kk}^{(k)}}}. \end{aligned}$$

Finally, defining

$$\rho_{ik}^{(k)} = \frac{\sigma_{ik}^{(k)}}{\sqrt{\sigma_{ii}^{(k)}\sigma_{kk}^{(k)}}}, \tag{A.17}$$

it is possible to compute

$$p_{ik} \,|\, i > k = \rho_{ik}^{(k)} \sqrt{\sigma_{ii}^{(k)}}. \tag{A.18}$$

Appendix 1.B Proposition 1

The demonstration that

$$MSE\left[\xi_t^{(i)}\left(H\right)\right]=\sum_{h=0}^{H-1}\Phi_h\Sigma\Phi_h'-\sigma_{ii}^{-1}\left(\sum_{h=0}^{H-1}\Phi_h\Sigma e_ie_i'\Sigma\Phi_h'\right). \tag{B.1}$$

Starting from

$$\xi_t^{(i)}\left(H\right)=\sum_{h=0}^{H-1}\Phi_h\left(u_{t+H-1-h}-\sigma_{ii}^{-1}\Sigma e_iu_{i,\,t+H-1-h}\right), \tag{B.2}$$

It can be verified that

$$E\left[\xi_t^{(i)}\left(H\right)\right]=0. \tag{B.3}$$

Then,

$$MSE\left[\xi_t^{(i)}\left(H\right)\right]=\text{var}\left[\sum_{h=0}^{H-1}\Phi_h\left(u_{t+H-1-h}-\sigma_{ii}^{-1}\Sigma e_iu_{i,\,t+H-1-h}\right)\right]. \tag{B.4}$$

As u_t is serially uncorrelated and for the properties of variance,

$$MSE\left[\xi_t^{(i)}\left(H\right)\right]=\sum_{h=0}^{H-1}\Phi_h\left\{a_h+b_h-2c_h\right\}\Phi_h', \tag{B.5}$$

where

$$a_h=\text{var}\left[u_{t+H-1-h}\right]=\Sigma, \tag{B.6}$$

$$\begin{aligned}b_h&=\text{var}\left[\sigma_{ii}^{-1}\Sigma e_iu_{i,\,t+H-1-h}\right]\\&=\sigma_{ii}^{-1}\Sigma_ue_i\sigma_{ii}e_i'\Sigma\sigma_{ii}\\&=\sigma_{ii}^{-1}\Sigma_ue_ie_i'\Sigma,\end{aligned} \tag{B.7}$$

and

$$
\begin{aligned}
c_h &= \text{cov}\left[u_{t+H-1-h},\, \sigma_{ii}^{-1}\Sigma e_i u_{i,\,t+H-1-h}\right] \\
&= \sigma_{ii}^{-1}\text{cov}\left[u_{t+H-1-h},\, e_i u_{i,\,t+H-1-h}\right]\Sigma \\
&= \sigma_{ii}^{-1}\Sigma e_i e_i' \Sigma. \qquad \text{(B.8)}
\end{aligned}
$$

Then,

$$
MSE\left[\xi_t^{(i)}(H)\right] = \sum_{h=0}^{H-1} \Phi_h \Sigma \Phi_h' - \sigma_{ii}^{-1}\left(\sum_{h=0}^{H-1} \Phi_h \Sigma e_i e_i' \Sigma \Phi_h'\right), \qquad \text{(B.9)}
$$

quod erat demonstrandum.

Appendix 1.C Proposition 2

The demonstration that

$$MSE\left[\xi_t^{(i)}\left(H\right)\right]=\sum_{h=0}^{H-1}\Phi_h\Sigma_t\left(H-h\right)\Phi_h^{\prime}$$
$$-\sigma_{ii,\,t}^{-1}\left(H-h\right)\left(\sum_{h=0}^{H-1}\Phi_h\Sigma_t\left(H-h\right)e_ie_i^{\prime}\Sigma_t\left(H-h\right)\Phi_h^{\prime}\right). \quad \text{(C.1)}$$

Starting from

$$\xi_t^{(i)}\left(H\right)=\sum_{h=0}^{H-1}\Phi_h\left(u_{t+H-h}-\sigma_{ii,\,t}^{-1}\left(H-h\right)\Sigma_t\left(H-h\right)e_iu_{i,\,t+H-h}\right), \quad \text{(C.2)}$$

such an approximation leads to

$$E\left[\xi_t^{(i)}\left(H\right)\right]=0. \quad \text{(C.3)}$$

Then

$$\text{MSE}\left[\xi_t^{(i)}\left(H\right)\right]=\text{var}\left[\sum_{h=0}^{H-1}\Phi_h\left(u_{t+H-h}-\sigma_{ii,\,t}^{-1}\left(H-h\right)\Sigma_t\left(H-h\right)e_iu_{i,\,t+H-h}\right)\right]. \quad \text{(C.4)}$$

As the residuals are the only random quantities in the variance argument and remembering that they are serially uncorrelated, it is possible to write:

$$\text{MSE}\left[\xi_t^{(i)}\left(H\right)\right]=\sum_{h=0}^{H-1}\text{var}\left[\Phi_h\left(u_{t+H-h}\right.\right.$$
$$\left.\left.-\sigma_{ii,\,t}^{-1}\left(H-h\right)\Sigma_t\left(H-h\right)e_iu_{i,\,t+H-h}\right)\right],$$

$$(C.5)$$

C.5 can be written as

$$MSE\left[\xi_t^{(i)}(H)\right]=\sum_{h=0}^{H-1}\mathit{\Phi}_h\left\{a_h+b_h-2c_h\right\}\mathit{\Phi}_h', \tag{C.6}$$

where

$$a_h=\text{var}\left[u_{t+H-h}\right]=\Sigma_t\left(H-h\right), \tag{C.7}$$

$$\begin{aligned}b_h&=\text{var}\left[\sigma_{ii,\,t}^{-1}\left(H-h\right)\Sigma_t\left(H-h\right)e_iu_{i,\,t+H-h}\right]\\&=\sigma_{ii,\,t}^{-1}\left(H-h\right)\Sigma_t\left(H-h\right)e_i\sigma_{ii,\,t}^{-1}\left(H-h\right)e_i'\\&\times\Sigma_t\left(H-h\right)\sigma_{ii,\,t}^{-1}\left(H-h\right)\\&=\sigma_{ii,\,t}^{-1}\left(H-h\right)\Sigma_t\left(H-h\right)e_ie_i'\Sigma_t\left(H-h\right),\end{aligned} \tag{C.8}$$

and

$$\begin{aligned}c_h&=\text{cov}\left[u_{t+H-h},\,\sigma_{ii,\,t}^{-1}\left(H-h\right)\Sigma_t\left(H-h\right)e_iu_{i,\,t+H-h}\right]\\&=\sigma_{ii,\,t}^{-1}\left(H-h\right)\text{cov}\left[u_{t+H-h},\,e_iu_{i,\,t+H-h}\right]\Sigma_t\left(H-h\right)\\&=\sigma_{ii,\,t}^{-1}\left(H-h\right)\\&\times\Sigma_t\left(H-h\right)e_ie_i'\Sigma_t\left(H-h\right).\end{aligned} \tag{C.9}$$

Then,

$$
\begin{aligned}
MSE\left[\xi_t^{(i)}(H)\right] &= \sum_{h=0}^{H-1} \varPhi_h \Sigma_t (H-h)\, \varPhi_h' \\
&- \sigma_{ii,t}^{-1} (H-h) \\
&\times \left(\sum_{h=0}^{H-1} \varPhi_h \Sigma_t (H-h)\, e_i e_i' \Sigma_t (H-h)\, \varPhi_h' \right),
\end{aligned} \tag{C.10}
$$

quod erat demonstrandum.

Appendix 1.D A Calculator Efficient Algorithm

The following formulation could be useful to optimize computations on calculators; it will be presented for the generalized heteroscedastic case as the homoscedastic and the orthogonal cases are just specialisations of the following formulas:

$$\mathrm{D\dot{Y}}_t^{(GVD)}(H) = \dot{S}_{1,t}^{(GVD)}(H) \circ \dot{S}_{2,t}^{(inv,GVD)}(H), \tag{D.1}$$

where $\mathrm{D\dot{Y}}_t^{(GVD)}(H)$ is the matrix containing the variance decompositions, the symbol $\circ$ identifies the Hadamard element by element product, and

$$\dot{S}_{1,t}^{(GVD)}(H) = \sum_{h=1}^{H-1} \Upsilon_h \dot{Q}_t^{(GVD)}(H)\, \Upsilon_h', \tag{D.2}$$

where Υ_h is a $(K \times KH)$ matrix such that $\Upsilon_h = [0 \ldots I_K \ldots 0]$ and the element in the first row and $[(h-1)K+1]$-th column is one, and

$$\dot{S}_{2,t}^{(inv,GVD)}(H) = \iota\iota' \left[\dot{S}_{2,t}^{(GVD)}(H)\right]^{-1}, \tag{D.3}$$

with ι being a $(K \times 1)$ vector of ones and

$$\dot{S}_{2,t}^{(GVD)}(H) = \sum_{h=1}^{H-1} \Upsilon_h \mathrm{diag}\left[\dot{Z}_t^{(GVD)}(H)\right] \Upsilon_h'. \tag{D.4}$$

Then,

$$\dot{Q}_t^{(GVD)}(H) = \left[\dot{D}_t^{(GVD)}(H)\right]^{-1} \dot{M}_t^{(GVD)}(H), \tag{D.5}$$

$$\dot{D}_t^{(GVD)}(H) = \oplus_{h=0}^{H-1} \dot{D}_t(H-h), \tag{D.6}$$

where $\dot{D}_t(H-h)$ is the forecast of the conditional diagonal variance matrix, and the symbol $\oplus$ identifies the operation of direct matrix sum,

$$\dot{M}_t^{(GVD)}(H) = \left[\Phi_t^{(GVD)}(H)\,\Sigma_t^{(GVD)}(H)\right] \circ \left[\Phi_t^{(GVD)}(H)\,\Sigma_t^{(GVD)}(H)\right]', \quad \text{(D.7)}$$

$$\Phi_t^{(GVD)}(H) = \oplus_{h=0}^{H-1}\mathit{\Phi}_h, \quad \text{(D.8)}$$

$$\Sigma_t^{(g)}(H) = \oplus_{h=0}^{H-1}\Sigma_t\,(H-h) \quad \text{(D.9)}$$

and

$$\dot{Z}_t^{(GVD)}(H) = \Phi_t^{(GVD)}(H)\,\Sigma_t^{(GVD)}(H)\left[\Phi_t^{(GVD)}(H)\right]'. \quad \text{(D.10)}$$

The advantage of such a formulation could be the fact that the matrices $\dot{Q}_t^{(GVD)}(H)$ and $\dot{Z}_t^{(GVD)}(H)$ are computed only once, as well as $\dot{M}_t^{(GVD)}(H)$, $\Phi_t^{(GVD)}(H)$ and $\Sigma_t^{(GVD)}(H)$. Then it is possible to store them and access them only when needed, with a saving in memory usage. Note also that the matrices $\dot{M}_t^{(GVD)}(H)$, $\Phi_t^{(GVD)}(H)$ and $\Sigma_t^{(GVD)}(H)$ can be erased from memory, as only $\dot{Q}_t^{(GVD)}(H)$ and $\dot{Z}_t^{(GVD)}(H)$ are needed for the final calculations. These features, as well as the generalized form of the Diebold Yilmaz Framework, could be particularly useful in high dimensional contexts.

The normalizations can then be performed respectively by

$$\widetilde{\dot{\mathrm{DY}}}_t^{(To,\,GVD)}(H) = \dot{\mathrm{DY}}_t^{(GVD)}(H) \circ \dot{S}_t^{(To,\,inv,\,GVD)}(H) \quad \text{(D.11)}$$

and

$$\widetilde{\dot{\mathrm{DY}}}_t^{(From,\,GVD)}(H) = \dot{\mathrm{DY}}_t^{(GVD)}(H) \circ \dot{S}_t^{(From,\,inv,\,GVD)}(H), \quad \text{(D.12)}$$

with

$$\dot{S}_t^{(To,\,inv,\,GVD)}\,(H):\;\dot{S}_t^{(To,\,inv,\,GVD)}\,(H)=\left[\mathbf{D}\dot{\mathbf{Y}}_t^{(GVD)}\iota_K\iota_K'\right]^{-1} \tag{D.13}$$

and

$$\dot{S}_t^{(From,\,inv,\,GVD)}\,(H):\;\dot{S}_t^{(From,\,inv,\,GVD)}\,(H)=\left[\iota_K\iota_K'\mathbf{D}\dot{\mathbf{Y}}_t^{(GVD)}\right]^{-1}. \tag{D.14}$$

Appendix 1.E A Method to Determine the Wold Order

A method to determine the Wold Order of a system based solely on the data could be the following.

Consider again the system

$$y_t = \nu + \sum_{l=1}^{p} A_l y_{t-l} + u_t, \tag{E.1}$$

where $y_t = [y_{1t},\, y_{2t},\, \ldots,\, y_{Kt}]$ and having variance covariance matrix Σ.

In order to compute orthogonal variance decompositions, one can apply the Upper Triangular Cholesky Decoposition and find a matrix P such that $P * P' = \Sigma$, so that $(P)^{-1}\,\Sigma\,(P')^{-1} = I_K$. The matrix P will then have the following form:

$$P = \begin{bmatrix} \sqrt{\sigma_{11}^{(1)}} & 0 & \dots & 0 \\ \sigma_{21}^{(1)}\left[\sigma_{11}^{(1)}\right]^{-1}\sqrt{\sigma_{11}^{(1)}} & \sqrt{\sigma_{22}^{(2)}} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{K1}^{(1)}\left[\sigma_{11}^{(1)}\right]^{-1}\sqrt{\sigma_{11}^{(1)}} & \sigma_{K2}^{(2)}\left[\sigma_{22}^{(2)}\right]^{-1}\sqrt{\sigma_{22}^{(2)}} & \dots & \sqrt{\sigma_{KK}^{(K)}} \end{bmatrix}, \tag{E.2}$$

where

$$\sigma_{ik}^{(j)} \mid (i < j \,\vee\, k < j) = 0, \tag{E.3}$$

$$\sigma_{ii}^{(j)} \,|\, i \leq j = \sigma_{ii}^{(i)}, \tag{E.4}$$

$$\sigma_{ik}^{(j)} \mid (i \geq j \,\wedge\, k \geq j) = \sigma_{ik}^{(j-1)} - \sigma_{i,\,j-1}^{(j-1)}\left[\sigma_{j-1,\,j-1}^{(j-1)}\right]^{-1}\sigma_{j-1,\,k}^{(j-1)}. \tag{E.5}$$

and can be used to transform the vectors of errors u_t into orthogonal vectors w_t in the following way:

$$w_t = P^{-1} u_t. \tag{E.6}$$

It turns out that (Hamilton [18], p.320) $w_{it}\sqrt{\sigma_{ii}^{(i)}}$ has the interpretation as the residual from a projection of u_{it} on $w_{1t}\sqrt{\sigma_{11}^{(1)}}, \ldots, w_{i-1,t}\sqrt{\sigma_{i-1,i-1}^{(i-1)}}$.

The goal is to find a method that maximises the explicative power of these K regressions. In this paper the following iterative algorithm is proposed:

1st Iteration: find the variable that has the most explicative power for the remaining $K-1$ variable by performing $K\,(K-1)$ univariate linear projections of every system variable on each one of the other variables and select the one that maximizes a criterion such as:

$$\gamma = t * R^2, \tag{E.7}$$

where t is the t-stat of the univariate parameter and R^2 is the centered coefficient of determination of each univariate regression.

Subsequent Iterations: at each iteration i, consider a new system that excludes the variable selected at the $(i-1)-th$ iteration and includes the residuals of the linear projections of all the other variables in the $(i-1)-th$ iteration on the selected variable of the $(i-1)-th$ iteration. Find, among the latter system, the variable that has the most explicative power for the remaining $K-i$ variable by performing $(K-i)\,(K-i-1)$ univariate linear projections of every system variable on each one of the other variables and select the one that maximizes the usual criterion:

$$\gamma = t * R^2. \tag{E.8}$$

If at a certain point in the iteration it is not possible to find any residual such that the inverse of its matrix product is not singular, the arlgorithm stops and the WO from that point on is arbitrary.

Table 1.1: **List of Variables Included in "Oxford-Man Institute's Realized Library"**

Series Name	Type of Data	Symbol	Series Number
S&P 500 (Live)	Realized Variance (5-minute)	SPX2.rv	1
FTSE 100 (Live)	Realized Variance (5-minute)	FTSE2.rv	2
Nikkei 225 (Live)	Realized Variance (5-minute)	N2252.rv	3
DAX (Live)	Realized Variance (5-minute)	GDAXI2.rv	4
Russel 2000 (Live)	Realized Variance (5-minute)	RUT2.rv	Not Used
All Ordinaries (Live)	Realized Variance (5-minute)	AORD2.rv	Not Used
DJIA (Live)	Realized Variance (5-minute)	DJI2.rv	Not Used
Nasdaq 100 (Live)	Realized Variance (5-minute)	IXIC2.rv	Not Used
CAC 40 (Live)	Realized Variance (5-minute)	FCHI2.rv	5
Hang Seng (Live)	Realized Variance (5-minute)	HSI2.rv	Not Used
KOSPI Composite Index (Live)	Realized Variance (5-minute)	KS11.rv	Not Used
AEX Index (Live)	Realized Variance (5-minute)	AEX.rv	Not Used
Swiss Market Index (Live)	Realized Variance (5-minute)	SSMI.rv	6
IBEX 35 (Live)	Realized Variance (5-minute)	IBEX2.rv	Not Used
S&P CNX Nifty (Live)	Realized Variance (5-minute)	NSEI.rv	Not Used
IPC Mexico (Live)	Realized Variance (5-minute)	MXX.rv	Not Used
Bovespa Index (Live)	Realized Variance (5-minute)	BVSP.rv	Not Used
Euro STOXX 50 (Live)	Realized Variance (5-minute)	STOXX50E.rv	Not Used
FT Straits Times Index	Realized Variance (5-minute)	FTSTI.rv	Not Used
FTSE MIB (Live)	Realized Variance (5-minute)	FTSEMIB.rv	Not Used

Table 1.2: **Full Sample Homoscedastic Orthogonal Connectedness Matrix**

The table shows OVDs obtained from a linear VAR estimated with full sample, along with some simple aggregations. The "TOTAL FROM" column reports the total variance received by each system equation from other variables, while the "TOTAL TO" row shows the total variance sent to other equations by the variable indexed by the column header. Finally, "TOTAL" indicates the total variance that is spilled across the system, which is equal to the sum of any one "TOTAL FROM", or "TOTAL TO" measures.

Series Number	1	2	3	4	5	6	TOTAL FROM
1	0.957	0.032	0.004	0.001	0.003	0.003	0.043
2	0.597	0.390	0.001	0.000	0.004	0.008	0.610
3	0.341	0.055	0.593	0.005	0.004	0.002	0.407
4	0.584	0.210	0.004	0.196	0.001	0.004	0.804
5	0.610	0.230	0.004	0.050	0.102	0.005	0.898
6	0.514	0.221	0.004	0.040	0.014	0.207	0.793
TOTAL TO	2.647	0.747	0.016	0.098	0.025	0.022	
TOTAL							59.254

Table 1.3: **Full Sample Heteroscedastic Orthogonal Connectedness Matrix**

The table shows OVDs constructed from a VAR with diagonal GARCH-DCC conditional heteroscedasticity. The "TOTAL FROM" column reports the total variance received by each system equation from other variables, while the "TOTAL TO" row shows the total variance sent to other equations by the variable indexed by the column header. Finally, "TOTAL" indicates the total variance that is spilled across the system, which is equal to the sum of any one of "TOTAL FROM", or "TOTAL TO" measures.

Series Number	1	2	3	4	5	6	TOTAL FROM
1	0.967	0.024	0.003	0.001	0.002	0.003	0.033
2	0.597	0.390	0.001	0.001	0.002	0.009	0.610
3	0.289	0.062	0.635	0.007	0.004	0.003	0.365
4	0.517	0.231	0.003	0.244	0.001	0.005	0.756
5	0.580	0.246	0.005	0.061	0.103	0.005	0.897
6	0.456	0.251	0.002	0.058	0.005	0.228	0.772
TOTAL TO	2.439	0.815	0.014	0.127	0.015	0.024	
TOTAL							57.224

Table 1.4: **Full Sample Homoscedastic Generalized Variance Decompositions Matrix**

The table shows GVDs constructed from a VAR with diagonal GARCH-DCC conditional heteroscedasticity. The "TOTAL FROM" column reports the total variance received by each system equation from other variables, while the "TOTAL TO" row shows the total variance sent to other equations by the variable indexed by the column header. Finally, "TOTAL" indicates the total variance that is spilled across the system; depending on which normalization scheme the analyst chooses, the total variance spilled is equal to the sum of either "TOTAL FROM", or "TOTAL TO" measures. In this case, since the measures were normalized by total variance spilled from each variable, the total is equal to the sum of the elements in the "TOTAL FROM" row.

Series Number	1	2	3	4	5	6	TOTAL TO
1	0.280	0.200	0.050	0.146	0.169	0.156	0.720
2	0.173	0.245	0.049	0.167	0.187	0.179	0.755
3	0.154	0.135	0.379	0.121	0.100	0.112	0.621
4	0.165	0.192	0.049	0.223	0.192	0.179	0.777
5	0.172	0.210	0.036	0.180	0.220	0.181	0.780
6	0.157	0.200	0.051	0.168	0.175	0.249	0.751
TOTAL FROM	0.821	0.938	0.235	0.782	0.823	0.807	
TOTAL							73.416

Table 1.5: **Full Sample Heteroscedastic Generalized Variance Decompositions Matrix**

The table shows GVDs constructed from a VAR with diagonal GARCH-DCC conditional heteroscedasticity. The "TOTAL FROM" column reports the total variance received by each system equation from other variables, while the "TOTAL TO" row shows the total variance sent to other equations by the variable indexed by the column header. Finally, "TOTAL" indicates the total variance that is spilled across the system; depending on which normalization scheme the analyst chooses, the total variance spilled is equal to the sum of either "TOTAL FROM", or "TOTAL TO" measures. In this case, since the measures were normalized by total variance spilled from each variable, the total is equal to the sum of the elements in the "TOTAL FROM" row.

Series Number	1	2	3	4	5	6	TOTAL TO
1	0.359	0.187	0.035	0.130	0.158	0.131	0.641
2	0.189	0.251	0.037	0.166	0.186	0.172	0.749
3	0.143	0.124	0.432	0.112	0.090	0.100	0.568
4	0.161	0.182	0.032	0.245	0.198	0.182	0.755
5	0.179	0.199	0.026	0.188	0.233	0.175	0.767
6	0.155	0.191	0.033	0.183	0.177	0.260	0.740
TOTAL FROM	0.827	0.883	0.164	0.778	0.809	0.759	
TOTAL							70.330

Table 1.6: **Percent Change from Homoscedastic to Heteroscedastic of Orthogonal Total From, Total To, and Grand Total Variance Decompositions**

The table reports the percent change in aggregate OVD based connectedness measures produced by the introduction of diagonal GARCH-DCC conditional heteroscedasticity in the VAR framework.

Series Number	ΔFROM	ΔTO	ΔTOTAL
1	-22.792%	-7.844%	
2	-0.007%	9.040%	
3	-10.216%	-14.670%	
4	-5.923%	30.316%	
5	-0.155%	-41.751%	
6	-2.685%	7.526%	
			-3.425%

Table 1.7: **Percentage Change from Homoscedastic to Heteroscedastic of Generalized Total From, Total To, and Grand Total Variance Decompositions**

The table reports the percent change in aggregate GVD based connectedness measures produced by the introduction of diagonal GARCH-DCC conditional heteroscedasticity in the VAR framework.

Series Number	ΔFROM	ΔTO	ΔTOTAL
1	0.744%	-10.976%	
2	-5.878%	-0.867%	
3	-30.304%	-8.568%	
4	-0.397%	-2.799%	
5	-1.723%	-1.644%	
6	-5.921%	-1.568%	
			-4.204%

Figure 1.1: **Homoscedastic and Heteroscedastic Orthogonal DY Measures**
The figure shows evolution of OVD based total connectedness obtained from rolling windows estimations of a linear VAR (in pink) and of a diagonal GARCH-DCC VAR (in black).

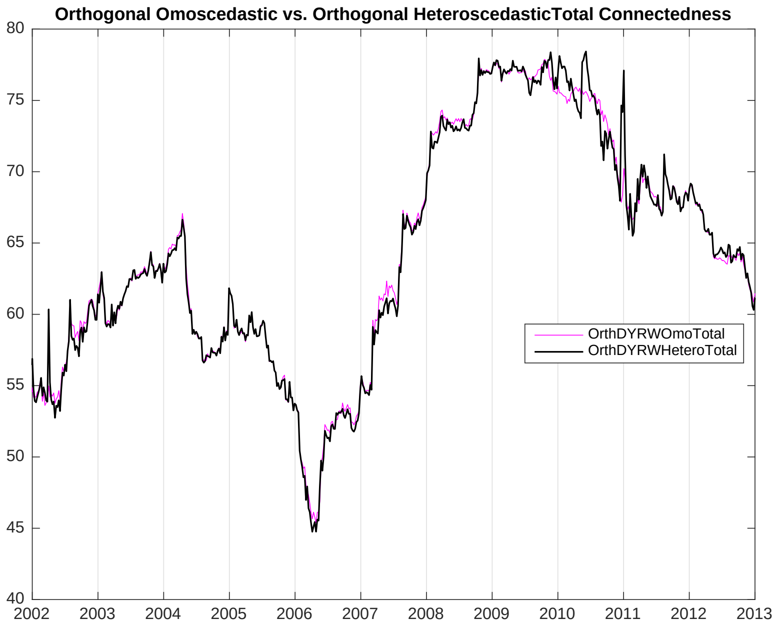


Figure 1.2: **Difference between Homoscedastic and Heteroscedastic Orthogonal DY Measures**
In this figure, the graph represents the percentage change in the rolling windows OVD based total connectedness, when a linear VAR is extended with a diagonal GARCH-DCC conditional heteroscedasticity.

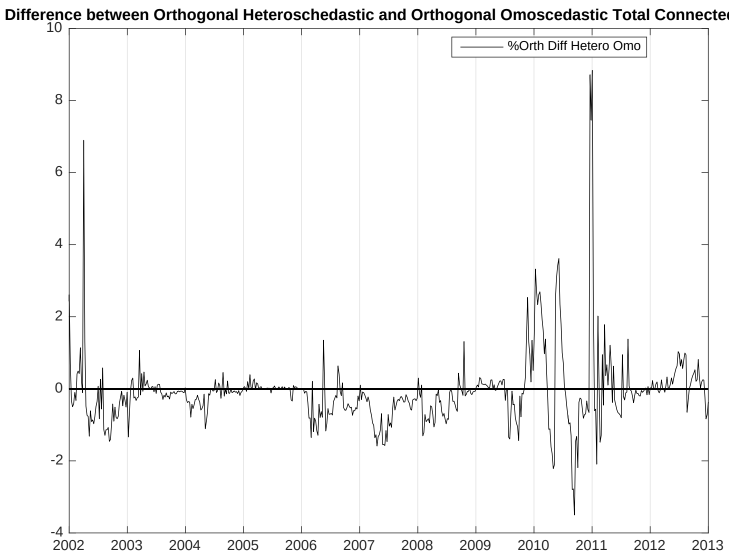


Figure 1.3: **Homoscedastic and Heteroscedastic Generalized DY Measures**
The figure shows evolution of GVD based total connectedness obtained from rolling windows estimations of a linear VAR (in pink) and of a diagonal GARCH-DCC VAR (in black).

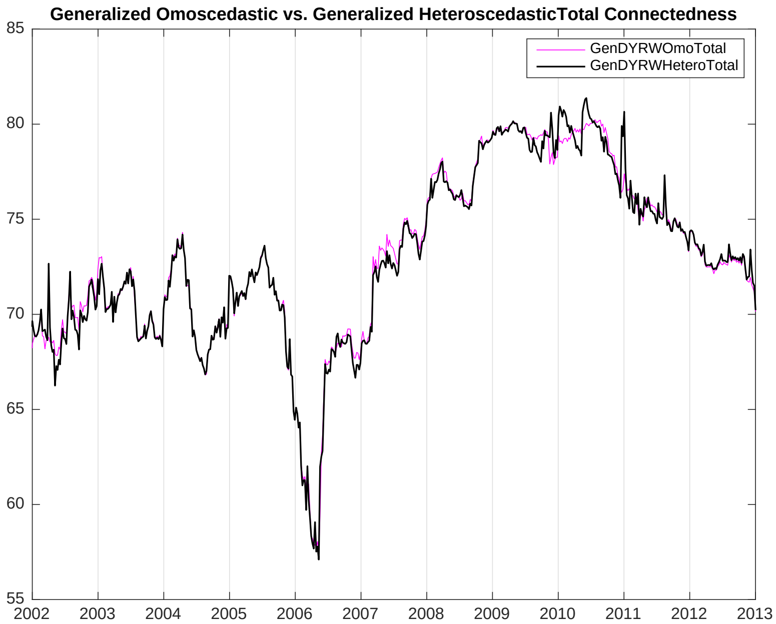


Figure 1.4: **Difference between Homoscedastic and Heteroscedastic Generalized DY Measures**
The figure shows the dynamic percent change in rolling windows GVD based total connectedness, when a linear VAR is extended with a diagonal GARCH-DCC conditional heteroscedasticity.

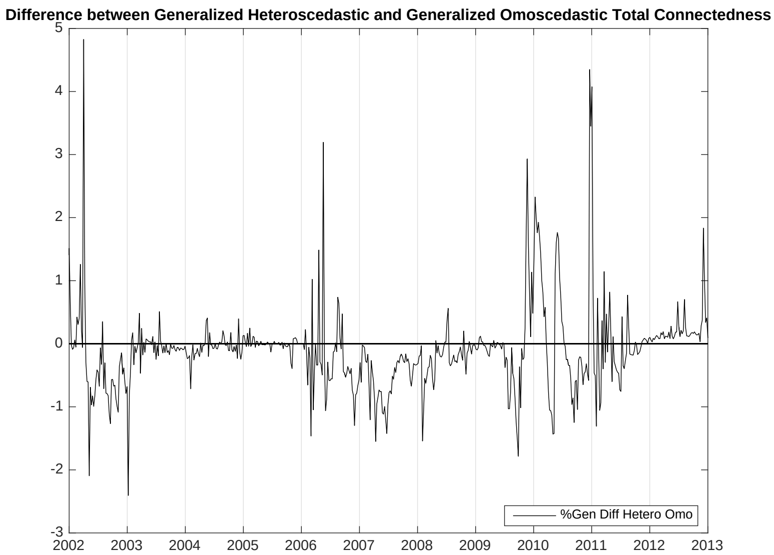
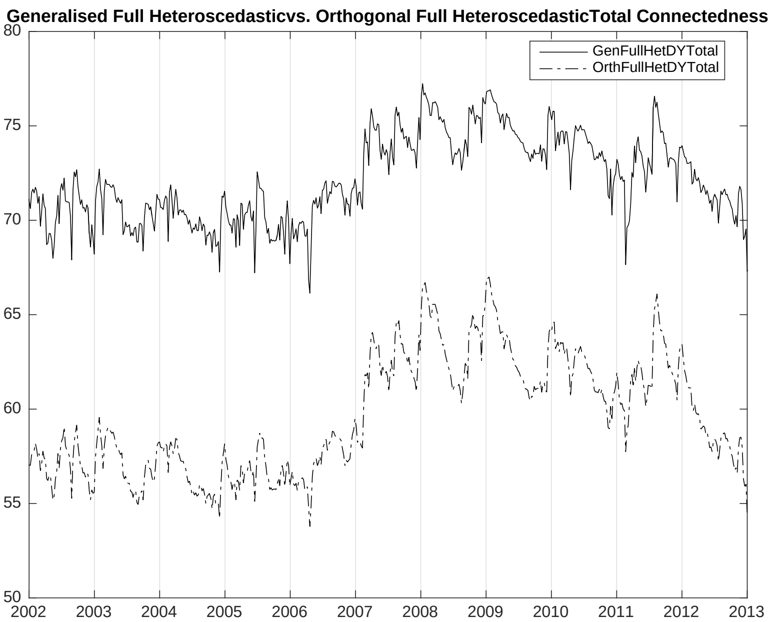


Figure 1.5: **Orthogonal and Generalized Smoothed Heteroscedastic DY Measures** This figure shows the evolution of total connectedness when the dynamics are being characterised only by conditional heteroscedasticity. The GVD based total connectedness is reported with bold black, while the OVD is reported as a dashed line.



Chapter 2

Generalized Diebold-Yilmaz Connectedness Measures for MS-VARs

Nicola Piazzalunga

September 13, 2018

Generalized Diebold-Yilmaz Connectedness Measures for MS-VARs

Abstract

The recent financial crisis has shown the world how important is to consider the dynamic nature of the strength of the relationships between financial variables. One way to assess both the magnitude and time variation of such strength is through the Diebold and Yilmaz connectedness framework, which is based on forecast error variance decompositions of VAR systems. Forecast error variance decompositions can be identified either through Cholesky decomposition, by assuming a prior Wold (or causal) Order, or via generalized variance decompositions, under some restrictions about the distribution of the system errors. The aim of this paper is to extend the Diebold and Yilmaz framework by relaxing such restrictions and proposing new formulae for the generalized variance decomposition of Markov switching vector autoregressions.

2.1 Introduction

One could argue that the economic crisis that begun around 2007 has been a highly multidimensional and nonlinear phenomenon. It would be hard to mention a subset of human interactions not touched by it, let alone the financial system. One of the issues raised in the discussions about the phenomena surrounding such a famous financial crisis was the striking amount of connectedness happening among an impressive amount of features of the financial sector (market risk, credit risk, portfolio management, pricing, systemic risk, etc.).

The Diebold and Yilmaz (DY) framework is one way to identify and analyze connectedness in a way that allows for some of both the high dimensionality and the nonlinearity. The DY framework was first proposed in Diebold and Yilmaz [9], further expanded in Diebold and Yilmaz [10], Diebold and Yılmaz [11] and consolidated in Diebold and Yilmaz [12]. The DY framework has been extensively applied to the study of network connectedness in economics and finance. Some recent applications are Diebold and Yilmaz [13], Bostanci and Yilmaz [5], Demirer et al. [7] and Diebold et al. [8].

The most simple specification of the DY framework has been outlined in Diebold and Yilmaz [9], which produces connectedness measures based on the variance decomposition of linear vector autoregressions (VAR). Each element of the variance decomposition forms a measure of variance (spillover) either directed from an element of the system to one of the others or from an element of the system to itself.

All these measures can be collected in a square matrix and form the primitives for the production of other measures of connectedness, such as, for example, the amount of variance directed from a subset of the system to another one, or the total connectedness of the system, identified as the amount of variance directed from (to) each element of the system to (from) the others. The most basic collection of connectedness measures is the connectedness table described forward in this paper, which also contains some of the measures mentioned above.

The variance decomposition is typically approximated by forecast variance decomposition (see, for example, Hamilton [20]). If the analyst has a prior about the Wold Order (WO, Wold [37]) of the VAR, the decomposition can be identified by Cholesky factorization as in Sims [35]. Typically, the analyst might not want to use a prior on the Wold order whenever she deals with highly chaotic relationships or with high dimensional systems. In such circumstances one could use generalized variance decompositions (GVDs) as in Pesaran and Shin [34], which do not require the specification of a WO, but "only" a tractable distributional assumption, typically normality.

The presence of many circumstances of second order with respect to the analyst focus, such as, for instance, non linearities or exogenous effects which are not modeled, might bring her to decide to allow the parameters of the relationship to vary. The DY framework accommodates for such a decision, being a dynamic analytical platform. One of the most direct way to induce parameter variation is to estimate the VAR through rolling windows. Such a method is a very general approximation of arbitrary nonlinear models, as Granger [17] pointed out, but the analyst might have reasons to make the time variation of the parameters more explicit, by specifying a different nonlinear model, for example for the motivations expressed in Ferrara et al. [14].

The focus of this paper is to extend the DY framework in order to explicitly accommodate for Markov switching vector autoregressions (MS- VAR) by proposing new formulae for the computation of generalized variance decompositions under the MS-VAR assumption.

The recent abundance of both data and computational power has made Markov Switching (MS) models an increasingly viable methodology for the study of a broad range of issues. Krolzig [27] collects the fundamental tools for the analysis of MS models, in particular in the context of Markov switching vector autoregressions. Originally, MS models were popularized in the econometric literature by Hamilton [19] as tools to assess abrupt variations in the parameter space of univariate autoregressions. The changes in the parameters can detect in fact a change in regime, such as a bear or bull market or a low versus high volatility regime. More formally, the parameter space varies conditional to a state variable following a Markov chain. According to Krolzig [27], the basic filtering and smoothing recursions to reconstruct the hidden Markov chain were introduced in Baum et al. [3] and applied by Lindgren [29] to regression models with Markovian regime switching. Kim [24] improved the smoother by backward recursion and Hansen [22] and Garcia [15] provided procedures for the estimation of the asymptotic null distribution useful for the determination of the number of regimes.

MS models are problems endowed with a high dimensional parameter space, especially when formulated as multivariate. A major contribution for the tractability of such models has come from handling them in a Bayesian framework. Albert and Chib [1] and McCulloch and Tsay [32] introduced the Gibbs sampling approach to the inference of MS models, while Carter and Kohn [6] formulated the problem in terms of multimove Gibbs sampling. The latter technique is particularly useful in the context of multivariate MSM, as Krolzig [27] points out, as it accelerates the convergence of the sampling algorithm. Finally, Sims et al. [36] provided methods for the inference of MSMs in the context of large multiple equation systems in the form of parameter restrictions.

The literature on MSMs is indeed vast and the small review above could only aim at roughly sketch it. More useful surveys can be found in Hamilton [21], Ang and Timmermann [2] and Guidolin et al. [18]. Recent examples of applications of MS-VARs can be found in Billio et al. [4] and Guidolin et al. [18], among others.

The structure of the paper is the following: section 2 provides the basics of a MS-VAR framework, section 3 develops explicit formulae for connectedness arising from a time variant autoregressive parameters (MSA) specification, section 4 generalizes to other specifications and section 5 will offer two experiments based on Montecarlo simulation. Finally, section 6 concludes the paper.

2.2 The MS-VAR framework

Markov switching models were introduced in the econometric literature to specify dynamics subject to sudden changes in parameters. Such models have been used to describe the behavior of both economic and financial time series, for example during financial panics, wars, or change in policies. A simple classification of such models distinguishes them according to which parameters vary: MSI referes to switching intercepts, while MSM points to switching means, MSA to switching autoregressive parameters and MSH to switching variance covariance matrices. Finally, we can model mixtures of those specifications up to the unrestricted MSIAH or MSMAH.

In the MS framework, VAR parameters are conditional to a state variable s_t, the dynamics of which are modeled as an M-states Markov chain of order one, that is:

$$P\left\{s_t = j|\, s_{t-1} = i,\, s_{t-2} = k,\, \ldots\right\} = P\left\{s_t = j|\, s_{t-1} = i\right\} = p_{ij}, \qquad (2.1)$$

where p_{ij} is called transition probability and $i,\, j = 1,\, 2,\, \ldots,\, M$, with M being the number of states.

As a consequence, it is possible to write a general MS-VAR(p) model as

$$E\left[y_t|\, Y_t,\, s_t\right] = \nu\left(s_t\right) + \sum_{l=1}^{p} A_l\left(s_t\right) y_{t-1}, \qquad (2.2)$$

$$u_t \equiv y_t - E\left[y_t|\, Y_t,\, s_t\right] \sim NID\left(0,\, \Sigma\left(s_t\right)\right), \qquad (2.3)$$

where $y_t \in \mathcal{M}\left(K \times 1\right)$ and $\nu\left(s_t\right)$, $A_l\left(s_t\right)$ and $\Sigma\left(s_t\right)$ are matrices of conditional parameters, Y_t represtents the sample available up to time t and u_t is a normally indipendently distributed error term.

Transition probabilities are collected in a $(M \times M)$ transition matrix such as

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{bmatrix}, \tag{2.4}$$

so that $\sum_{m=1}^{M} p_{im} = 1$, as in Krolzig [27].

It is convenient to characterize the Markov chain by defining a state vector as

$$\xi_t = \begin{bmatrix} \delta(s_t = 1) \\ \delta(s_t = 2) \\ \vdots \\ \delta(s_t = M) \end{bmatrix}, \tag{2.5}$$

where $\delta(\cdot)$ is the indicator function. Naturally, $\sum_{m}^{M} \xi_{mt} = 1$. Moreover,

$$E[\xi_t | s_t = i] = P' e_i, \tag{2.6}$$

where e_i is the i-th column of the identity matrix of order M, and

$$E[\xi_t] = \begin{bmatrix} Pr(\xi_t = e_1) \\ Pr(\xi_t = e_2) \\ \vdots \\ Pr(\xi_t = e_M) \end{bmatrix}. \tag{2.7}$$

Then, the dynamics of ξ_t can be described as a $VAR(1)$ such as in

$$\xi_{t+1} = P' \xi_t + v_t, \tag{2.8}$$

where

$$v_t \equiv \xi_{t+1} - E\left[\xi_{t+1}\middle|\, \xi_t,\, \xi_{t-1},\, \ldots\right] \tag{2.9}$$

is a martingale difference sequence (MDS), which can display only a finite set of values, but is zero on average and cannot be forecasted on the basis of previous states of the process.

The state vector itself can be forecasted as

$$E\left[\xi_{t+T}\middle|\, \xi_t,\, \xi_{t-1},\, \ldots\right] = \left(P'\right)^H \xi_t, \tag{2.10}$$

so that

$$\begin{bmatrix} Pr\left(s_{t+H} = 1|s_t = i\right) \\ Pr\left(s_{t+H} = 2|s_t = i\right) \\ \vdots \\ Pr\left(s_{t+H} = M|s_t = i\right) \end{bmatrix} = \left(P'\right)^H e_i. \tag{2.11}$$

Finally, if one eigenvalue of P is one and all the others are inside the unit circle, the Markov chain is ergodic and it is possible to define a vector of ergodic probabilities π such that

$$P'\pi = \pi \tag{2.12}$$

and

$$\lim_{T\to\infty} \left(P'\right)^T = \pi 1'_M, \tag{2.13}$$

where 1_M is a $(M \times 1)$ vector, which elements are all equal to one.

It can be shown that an ergodic Markov chain is covariance stationary. Ergodic probabilities can be obtained as

$$\pi = (W'W)^{-1} W' e_{M+1}, \tag{2.14}$$

where

$$W = \begin{bmatrix} I_M - P' \\ 1'_K \end{bmatrix}, \tag{2.15}$$

e_{M+1} is the $M+1$-th column of the identity matrix I_{M+1}.

The formulation of a MS-VAR state space critically depends on whether its mean is state dependent. If the model can be expressed as having a state dependent intercept (MSI-VAR), the formulation can be a declination of formulae 2.2 and 2.3. On the other hand, when the MS-VAR is expressed as departures from the state dependent means (MSM-VAR), then the conditional density of y_t depends on the last $p+1$ regimes and the Markov chain must be adjusted as follows.

First of all, the state vector is extended in order to take into account the dependency on the last regimes as

$$\xi_t = \xi_t^{(p+1)} = \otimes_{l=0}^{p} \xi_{t-l}^{(1)} = \xi_t^{(1)} \otimes \xi_{t-2}^{(1)} \otimes \ldots \otimes \xi_{t-p}^{(1)}, \tag{2.16}$$

so that the number of states, called derived states, is now $N = M^{(p+1)}$ and where $\xi_t^{(1)}$ is the primitive state vector. As a consequence, ξ_t is defined, for MSM processes, as

$$\xi_t = \xi_t^{(p+1)} = \begin{bmatrix} \delta\left(s_{t-1} = 1,\, s_{t-2} = 1,\, \ldots,\, s_{t-p} = 1\right) \\ \delta\left(s_{t-1} = 1,\, s_{t-2} = 1,\, \ldots,\, s_{t-p} = 2\right) \\ \vdots \\ \delta\left(s_{t-1} = M,\, s_{t-2} = M,\, \ldots,\, s_{t-p} = M\right) \end{bmatrix}, \tag{2.17}$$

while it can be shown that

$$\xi_t^{(r)} = \left(I_{M^r} \otimes 1'_{M^{p+1-r}} \right) \xi_t^{p+1},\ r \leq p+1, \tag{2.18}$$

and

$$\xi_{t-l}^{(1)} = \left(1'_{M^l} \otimes I_M \otimes 1'_{M^{p-l}} \right) \xi_t^{(p+1)}. \tag{2.19}$$

In order to express the derived state vector as a vector autoregression, a derived transition matrix F must be created. Appendix 2.A shows that F should satisfy

$$\begin{aligned} Pr\left(\xi_{t+1} = \xi_{t+1}^{*} |\, \xi_t\right) = Pr&\left(\xi_{t+1}^{(1)} = \xi_{t+1}^{(1)*} |\, \xi_t^{(1)}\right) \\ &\times Pr\left(\xi_t^{(1)} = \xi_t^{(1)*},\, \xi_{t-1}^{(1)} = \xi_{t-1}^{(1)*},\, \ldots,\, \xi_{t-p+1}^{(1)} = \xi_{t-p+1}^{(1)*}\right. \\ &\left. |\, \xi_t^{(1)},\, \xi_{t-1}^{(1)},\, \ldots,\, \xi_{t-p+1}^{(1)}\right), \end{aligned} \tag{2.20}$$

where the star superscript indicates a realization of the variable. In matrix form,

$$\begin{aligned} E\left[\xi_{t+1} |\, \xi_t\right] &= \left\{\text{diag}\left[\text{vec}\left(P\right) \otimes 1_{M^{p-1}}\right]\right\} \left(1_M \otimes \xi_t^{(p)}\right) \\ &= \left\{\text{diag}\left[\text{vec}\left(P\right) \otimes 1_{M^{p-1}}\right]\right\} \left(1_M \otimes I_{M^p} \otimes 1'_M\right) \xi_t, \end{aligned} \tag{2.21}$$

so that

$$F = \left\{\text{diag}\left[\text{vec}\left(P\right) 1_{M^{p-1}}\right]\right\} \left(1_M \otimes I_{M^p} \otimes 1'_M\right). \tag{2.22}$$

Finally, the Markov chain can be now written as

$$\xi_{t+1} = F \xi_{t-1} + v_{t+1}. \tag{2.23}$$

Although a full review of the estimation of MS-VARs is clearly outside the scope of this paper, we report in the Appendix 2.B a method for the estimation of MSM-VARs that uses the EM algorithm. First, the estimation technique is introduced for a MSM formulation, then it will be expanded to cover the most general MSMAH, which is a fully unrestricted model. A simplified version of such methods is used for the estimation of MSI-VARs, so it will be omitted. Readers can always refer to Krolzig [27] for a comprehensive coverage on the matter.

2.3 The connectedness framework

Diebold and Yilmaz [9] and Diebold and Yilmaz [10] introduced and extended the connectedness table as a $(K+1)\times(K+1)$ matrix, which elements are $(K \times K)$ variance decompositions, and $2K+1$ aggregate measures: K variance spilled to other variables, K variances spilled from other variables and a total variance spilled across the system. Depending on the method of variance decomposition, one can have an orthogonal or a generalized table.

Orthogonal Connectedness Table The orthogonal connectedness table (OCT) was introduced in Diebold and Yilmaz [9] as a way to collect the most basic connectedness measures. It is based on forecast errors orthogonal variance decompositions (OVDs) obtained through a Cholesky decomposition of the variance covariance matrix of the VAR error term. As such, the OCT is dependent on a specific WO defined a priori by the analyst. The table is structured as follows, with the total directional connectedness measures from other variables in the $K+1$-th column and the total directional connectedness measures to other variables in the $K+1$-th row:

$$
\begin{array}{cccccc}
 & y_1 & y_2 & \dots & y_K & From\,Others \\
 & & & & & \\
y_1 & d^H_{11} & d^H_{12} & \dots & d^H_{1K} & \sum_{j=1}^K d^H_{1j},\, j \neq 1 \\
y_2 & d^H_{21} & d^H_{22} & \dots & d^H_{2K} & \sum_{j=1}^K d^H_{2j},\, j \neq 2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
y_K & d^H_{K1} & d^H_{K2} & \dots & d^H_{KK} & \sum_{j=1}^K d^H_{Kj},\, j \neq K \\
 & & & & & Total \\
 & \sum_{i=1}^K d^H_{i1}, & \sum_{i=1}^K d^H_{i1},\, i \neq 1 & \dots & \sum_{i=1}^K d^H_{i1},\, i \neq 1 & \sum_{i,\,j=1}^K d^H_{ij} \\
To\,Others & i \neq 1 & i \neq 2 & & i \neq K & i \neq j
\end{array}
\tag{2.24}
$$

Generalized Connectedness Table The generalized connectedness table is based on forecast errors generalized variance decompositions (GVDs), as described in the seminal contribution by Pesaran and Shin [34] and applied to the connectedness framework first in Diebold and Yilmaz [10]. Because of the way the matrix is computed, the total directional connectedness measures to other variables are contained in the $K + 1$-th column, while the total directional connectedness measures from others are reported in the $K + 1$-th row. This paper focuses on the development of closed form formulae for the computation of the building blocks of generalized connectedness tables, an instance of which is displayed below:

$$
\begin{array}{cccccc}
 & y_1 & y_2 & \dots & y_K & To\,Others \\
 & & & & & \\
y_1 & d^H_{11} & d^H_{12} & \dots & d^H_{1K} & \sum_{j=1}^K d^H_{1j},\, j \neq 1 \\
y_2 & d^H_{21} & d^H_{22} & \dots & d^H_{2K} & \sum_{j=1}^K d^H_{2j},\, j \neq 2 \\
\vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\
y_K & d^H_{K1} & d^H_{K2} & \dots & d^H_{KK} & \sum_{j=1}^K d^H_{Kj},\, j \neq K \\
 & & & & & Total \\
From\,Others & \sum_{i=1}^K d^H_{i1}, & \sum_{i=1}^K d^H_{i1},\, i \neq 1 & \dots & \sum_{i=1}^K d^H_{i1},\, i \neq 1 & \sum_{i,\,j=1}^K d^H_{ij} \\
 & i \neq 1 & i \neq 2 & & i \neq K & i \neq j
\end{array}
\tag{2.25}
$$

2.3.1 The Generalized Diebold Yilmaz measure of connectedness for the MSA(K)-VAR(1)

In a VAR setting, DY measures of connectedness are functions of only autoregressive parameters and the variance covariance matrix; that is, the DYMs do not depend on the intercept nor on the mean. For this reason, the paper will begin the analysis with the MSA case and will show that the property will be maintained in other Markov switching settings, such as MSIA and MSMA.

Consider the MSA (K) -VAR (1) written in its state-space form

$$\begin{aligned}
y_t &= A\left(s_t\right) y_{t-1} + \varepsilon_t \\
\xi_t &= F\xi_{t-1} + v_t, \\
F &= P'.
\end{aligned} \tag{2.26}$$

In models with regime dependent autoregressive dynamics, the lagged endogenous variables y_{t-1} are likely correlated with the regime vector ξ_t, thus a difficulty arises in forecasting such MSMs. To mitigate this issue, following Krolzig et al. [28], the system is represented through the Karlsen linear state space representation (Karlsen [23]) $\psi_t = \xi_t \otimes y_t$; for $p = 1$, this can be written as

$$\begin{aligned}
\psi_t &= \xi_t \otimes y_t \\
&= \left[F\xi_{t-1} + v_t\right] \otimes \left[A\left(s_t\right) y_{t-1} + \varepsilon_t\right] \\
&= \left(F\xi_{t-1}\right) \otimes \left(A\left(s_t\right) y_{t-1}\right) + v_t \otimes \left(A\left(s_t\right) y_{t-1}\right) + \left(F\xi_{t-1}\right) \otimes \varepsilon_t + v_t \otimes \varepsilon_t \\
&= \left(F \otimes A\left(s_t\right)\right)\left(\xi_{t-1} \otimes y_{t-1}\right) + v_t \otimes \left(A\left(s_t\right) y_{t-1}\right) + \left(F\xi_{t-1} + v_t\right) \otimes \varepsilon_t \\
&= \Pi\psi_{t-1} + \epsilon_{t,\, s_t}
\end{aligned} \tag{2.27}$$

where

$$\epsilon_{t,\,s_t} = v_t \otimes \left(A\left(s_t\right)y_{t-1}\right) + \left(F\xi_{t-1} + v_t\right) \otimes \varepsilon_t \tag{2.28}$$

is a martingale difference sequence (will be simply referenced as ϵ_t to ease the notational burden). Moreover, we show in Appendix 2.C that

$$\Pi = F \otimes A\left(s_t\right) = \begin{bmatrix} p_{11}A_1 & p_{21}A_1 & \dots & p_{M1}A_1 \\ p_{12}A_2 & p_{22}A_2 & \dots & p_{M2}A_2 \\ \vdots & \vdots & \ddots & \vdots \\ p_{1M}A_M & p_{2M}A_M & \dots & p_{MM}A_M \end{bmatrix}, \tag{2.29}$$

so that a moving average representation of ψ can be written in the following form:

$$\psi_{t+H} = \Pi^H \psi_t + \sum_{h=0}^{H-1} \Pi^h \epsilon_{t+H-h}. \tag{2.30}$$

The forecast and the forecast error of the linear state space representation can be thus expressed as

$$E_t\left[\psi_{t+H}\right] = \Pi^H \psi_t \tag{2.31}$$

and

$$\begin{aligned} \omega^{(\psi)}\left(H\right) &= \psi_{t+H} - E_t\left[\psi_{t+H}\right] \\ &= \sum_{h=0}^{H-1} \Pi^h \epsilon_{t+H-h} \\ &= \sum_{h=0}^{H-1} \Pi^h \left[v_{t+H-h} \otimes \left(A\left(s_t\right)y_{t+H-h-1}\right)\right. \\ &\quad \left. + \left(F\xi_{t+H-h-1} + v_{t+H-h}\right) \otimes \varepsilon_{t+H-h}\right], \end{aligned} \tag{2.32}$$

so that the moving average representation, the forecast and the forecast error of y can be written as

$$\begin{aligned} y_{t+H} &= \sum_{k=1}^{K} \xi_{k,\,t+H} y_{t+H} \\ &= \left[1'_M \otimes I_K\right] \psi_{t+H} \\ &= \left[1'_M \otimes I_K\right] \left[\Pi^H \psi_t + \sum_{h=0}^{H-1} \Pi^h \epsilon_{t+H-h}\right], \end{aligned} \tag{2.33}$$

$$\begin{aligned} E_t\left[y_{t+H}\right] &= \sum_{k=1}^{K} E_t\left[\xi_{k,\,t+H} y_{t+H}\right] \\ &= \left[1'_M \otimes I_K\right] E_t\left[\psi_{t+H}\right] \\ &= \left[1'_M \otimes I_K\right] \Pi^H \psi_t \end{aligned} \tag{2.34}$$

and

$$\begin{aligned} \omega_t^{(y)}(H) &= y_{t+H} - E_t\left[y_{t+H}\right] \\ &= \left[1'_M \otimes I_K\right] \omega_t^{(\psi)}(H) \\ &= \left[1'_M \otimes I_K\right] \left\{ \sum_{h=0}^{H-1} \Pi^h \left[v_{t+H-h} \otimes \left(A\left(s_{t+H-h}\right) y_{t+H-h-1}\right)\right.\right. \\ &\quad + \left.\left.\left(F\xi_{t+H-h-1} + v_{t+h}\right) \otimes \varepsilon_{t+H-h}\right]\right\} \\ &= \left[1'_M \otimes I_K\right] \left\{ \left\{ \sum_{h=0}^{H-1} \Pi^h \left[v_{t+h} \otimes \left(A\left(s_{t+H-h}\right) y_{t+H-h-1}\right)\right] \right.\right. \\ &\quad \left. + \sum_{h=0}^{H-1} \Pi^h \left[v_{t+H-h} \otimes \varepsilon_{t+H-h}\right] \right\} \\ &\quad \left. + \sum_{h=1}^{H} \Pi^h \left[\left(F\xi_{t+H-h-1}\right) \otimes \varepsilon_{t+H-h}\right] \right\} \end{aligned}$$

$$= \omega_t^{(y,\varepsilon,v)}(H) + \omega_t^{(y,\varepsilon)}(H), \tag{2.35}$$

where

$$\omega_t^{(y,\varepsilon,v)}(H) = [1'_M \otimes I_K] \left\{ \sum_{h=0}^{H-1} \Pi^h \left[v_{t+h} \otimes \left(A\left(s_{t+H-h}\right) y_{t+H-h-1} \right) \right] \right.$$
$$\left. + \sum_{h=0}^{H-1} \Pi^h \left[v_{t+H-h} \otimes \varepsilon_{t+H-h} \right] \right\}, \tag{2.36}$$

and

$$\omega_t^{(y,\varepsilon)}(H) = [1'_M \otimes I_K] \left\{ \sum_{h=1}^{H} \Pi^h \left[(F\xi_{t+H-h-1}) \otimes \varepsilon_{t+H-h} \right] \right\}. \tag{2.37}$$

Since it is not yet possible to implement GVDs without assuming normality in the error process, this study will focus on the component of the total forecasting error for which the Gaussian innovation is the sole responsible, i.e. $\omega_t^{(y,\varepsilon)}(H)$.

Since the error term of the Markov chain has already been ruled out, the prediction of the state probabilities vector can be approximated with

$$\xi_{t+h} \approx \hat{\xi}_{t+h|t} = F^h \hat{\xi}_{t|t}, \tag{2.38}$$

so that the covariance of the normal component of the forecast error can be expressed as

$$\begin{aligned}
\operatorname{cov}\left(\omega_t^{(y,\varepsilon)}(H)\right) &= E_t\left[\omega_t^{(y,\varepsilon)}(H)\left(\omega_t^{(y,\varepsilon)}(H)\right)'\right] \\
&= E_t\left\{[1'_M \otimes I_K]\left\{\sum_{h=0}^{H-1}\Pi^h\left[\left(F\hat{\xi}_{t+H-h-1}\right)\otimes\varepsilon_{t+H-h}\right]\right.\right. \\
&\times \left.\left.\left[\left(F\hat{\xi}_{t+H-h-1}\right)\otimes\varepsilon_{t+H-h}\right]'\left(\Pi^h\right)'\right\}[1'_M \otimes I_K]'\right\}
\end{aligned}$$

$$
\begin{aligned}
&= \left[1_M' \otimes I_K\right] E_t \left\{ \sum_{h=0}^{H-1} \Pi^h \left[\left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \varepsilon_{t+H-h} \right] \right. \\
&\times \left. \left[\left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \varepsilon_{t+H-h} \right]' \left(\Pi^h\right)' \right\} \left[1_M' \otimes I_K\right]' \\
&= \left[1_M' \otimes I_K\right] \left\{ \sum_{h=1}^{H} \Pi^h E_t \left\{ \left[\left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \varepsilon_{t+H-h} \right] \right. \right. \\
&\times \left. \left. \left[\left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \varepsilon_{t+H-h} \right]' \right\} \left(\Pi^h\right)' \right\} \left[1_M' \otimes I_K\right]' \\
&= \left[1_M' \otimes I_K\right] \left\{ \sum_{h=0}^{H-1} \Pi^h E_t \left[\left[\left(F^{H-h} \hat{\xi}_{t|\,t} \right) \left(F^{H-h} \hat{\xi}_{t|\,t} \right)' \right] \right. \right. \\
&\otimes \left. \left. \left(\varepsilon_{t+H-h} \varepsilon_{t+H-h}' \right) \right] \left(\Pi^h\right)' \right\} \left[1_M' \otimes I_K\right]' \\
&= \left[1_M' \otimes I_K\right] \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|\,t} \right)^2 \right. \right. \\
&\otimes E_t \left. \left. \left[\left(\varepsilon_{t+H-h} \varepsilon_{t+H-h}' \right) \right] \right\} \left(\Pi^h\right)' \right\} \left[1_M' \otimes I_K\right]' \\
&= \left[1_M' \otimes I_K\right] \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|\,t} \right)^2 \otimes \Sigma \right\} \left(\Pi^h\right)' \right\} \left[1_M' \otimes I_K\right]' .
\end{aligned}
\tag{2.39}
$$

An application of Pesaran and Shin [34] leads to computing the forecast error conditional on information on the variable i as

$$
\begin{aligned}
\omega_t^{(y,\,\varepsilon,\,i)}(H) = \left[1_M' \otimes I_K\right] &\left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|\,t} \right) \right. \right. \\
&\otimes \left. \left. \left[\varepsilon_{t+H-h} - E\left(\varepsilon_{t+H-h} \middle|\, \varepsilon_{i,\,t+H-h} \right) \right] \right\} \right\} .
\end{aligned}
\tag{2.40}
$$

Assuming $\varepsilon_{t+H-h} \sim \mathcal{N}(0,\, \Sigma)$, the expectation in the equation above can be computed as

$$
E\left(\varepsilon_{t+H-h} \middle|\, \varepsilon_{i,\,t+H-h} \right) = \left(\sigma_{ii}^{-1} \Sigma e_i \right) \varepsilon_{i,\,t+H-h}, \tag{2.41}
$$

leading to

$$\omega_t^{(y,\,\varepsilon,\,i)}(H) = [1'_M \otimes I_K] \\ \times \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \left[\varepsilon_{t+H-h} - \left(\sigma_{ii}^{-1} \Sigma e_i \right) \varepsilon_{i,\,t+H-h} \right] \right\} \right\}. \quad (2.42)$$

so that the covariance matrix of the normal component of the forecast error conditional to the information on the i-th variable can be computed as

$$\begin{aligned} \operatorname{cov}\left(\omega_t^{(y,\,\varepsilon,\,i)}(H)\right) &= E_t\left[\omega_t^{(y,\,\varepsilon,\,i)}(H)\left(\omega_t^{(y,\,\varepsilon,\,i)}(H)\right)'\right] \\ &= [1'_M \otimes I_K]\, E_t \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|\,t} \right) \right. \right. \\ &\otimes \left[\varepsilon_{t+H-h} - \left(\sigma_{ii}^{-1} \Sigma e_i \right) \varepsilon_{i,\,t+H-h} \right] \Big\} \\ &\times \left\{ \left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \left[\varepsilon_{t+H-h} - \left(\sigma_{ii}^{-1} \Sigma e_i \right) \varepsilon_{i,\,t+H-h} \right] \right\}' \left(\Pi^h\right)' \Big\} \\ &\times [1'_M \otimes I_K]' \\ &= [1'_M \otimes I_K]\, E_t \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \varepsilon_{t+H-h} \right. \right. \\ &- \left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \left[\left(\sigma_{ii}^{-1} \Sigma e_i \right) \varepsilon_{i,\,t+H-h} \right] \Big\} \\ &\times \left\{ \left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \varepsilon_{t+H-h} \right. \\ &- \left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \left[\left(\sigma_{ii}^{-1} \Sigma e_i \right) \varepsilon_{i,\,t+H-h} \right] \Big\}' \left(\Pi^h\right)' \Big\} [1'_M \otimes I_K]' \\ &= [1'_M \otimes I_K] \left\{ \sum_{h=0}^{H-1} \Pi^h \left[a_h + b_h - 2c_h \right] \left(\Pi^h\right)' \right\} [1'_M \otimes I_K]', \end{aligned} \quad (2.43)$$

where

$$
\begin{aligned}
a_h &= E_t\left\{\left[\left(F^h\hat{\xi}_{t|t}\right)\otimes\varepsilon_{t+H-h}\right]\left[\left(F^{H-h}\hat{\xi}_{t|t}\right)\otimes\varepsilon_{t+H-h}\right]'\right\}\\
&= \left[\left(F^{H-h}\hat{\xi}_{t|t}\right)\left(F^{H-h}\hat{\xi}_{t|t}\right)'\right]\otimes E_t\left[\varepsilon_{t+H-h}\varepsilon'_{t+H-h}\right]\\
&= \left(F^{H-h}\hat{\xi}_{t|t}\right)^2\otimes\Sigma, \qquad (2.44)
\end{aligned}
$$

$$
\begin{aligned}
b_h &= E_t\left\{\left[\left(F^{H-h}\hat{\xi}_{t|t}\right)\otimes\left(\sigma_{ii}^{-1}\Sigma e_i\right)\varepsilon_{i,\,t+H-h}\right]\left[\left(F^{H-h}\hat{\xi}_{t|t}\right)\otimes\left(\sigma_{ii}^{-1}\Sigma e_i\right)\varepsilon_{i,\,t+H-h}\right]'\right\}\\
&= \left[\left(F^{H-h}\hat{\xi}_{t|t}\right)\left(F^{H-h}\hat{\xi}_{t|t}\right)'\right]\otimes\left\{\left(\sigma_{ii}^{-1}\Sigma e_i\right)E_t\left[\varepsilon_{i,\,t+H-h}\varepsilon'_{i,\,t+H-h}\right]\left(\sigma_{ii}^{-1}\Sigma e_i\right)'\right\}\\
&= \left(F^{H-h}\hat{\xi}_{t|t}\right)^2\otimes\left(\sigma_{ii}^{-1}\Sigma e_i\sigma_{ii}e_i'\Sigma'\sigma_{ii}^{-1}\right)\\
&= \sigma_{ii}^{-1}\left(F^{H-h}\hat{\xi}_{t|t}\right)^2\otimes\left(\Sigma e_i e_i'\Sigma'\right)\\
&= \sigma_{ii}^{-1}\left(F^{H-h}\hat{\xi}_{t|t}\right)^2\otimes\left(\Sigma e_i\right)^2 \qquad (2.45)
\end{aligned}
$$

and

$$
\begin{aligned}
c_h &= E_t\left\{\left[\left(F^{H-h}\hat{\xi}_{t|t}\right)\otimes\varepsilon_{t+H-h}\right]\left[\left(F^{H-h}\hat{\xi}_{t|t}\right)\otimes\left(\sigma_{ii}^{-1}\Sigma e_i\right)\varepsilon_{i,\,t+H-h}\right]'\right\}\\
&= \left[\left(F^{H-h}\hat{\xi}_{t|t}\right)\left(F^{H-h}\hat{\xi}_{t|t}\right)'\right]\otimes E_t\left\{\varepsilon_{t+H-h}\left[\left(\sigma_{ii}^{-1}\Sigma e_i\right)\varepsilon_{i,\,t+H-h}\right]'\right\}\\
&= \left(F^{H-h}\hat{\xi}_{t|t}\right)^2\otimes E_t\left\{\varepsilon_{t+H-h}\varepsilon'_{i,\,t+H-h}\sigma_{ii}^{-1}e_i'\Sigma'\right\}\\
&= \left(F^{H-h}\hat{\xi}_{t|t}\right)^2\otimes E_t\left\{\varepsilon_{t+H-h}\varepsilon'_{t+H-h}e_i\sigma_{ii}^{-1}e_i'\Sigma'\right\}\\
&= \left(F^{H-h}\hat{\xi}_{t|t}\right)^2\otimes\left(\sigma_{ii}^{-1}\Sigma e_i e_i'\Sigma'\right)\\
&= \sigma_{ii}^{-1}\left(F^{H-h}\hat{\xi}_{t|t}\right)^2\otimes\left(\Sigma e_i\right)^2. \qquad (2.46)
\end{aligned}
$$

Then, the conditional covariance matrix can be expressed as

$$
\begin{aligned}
\text{cov}\left(\omega_t^{(y,\varepsilon,\,i)}\left(H\right)\right) &= \left[1_M' \otimes I_K\right]\left\{\sum_{h=0}^{H-1}\Pi^h\left[a_h+b_h-2c_h\right]\left(\Pi^h\right)'\right\}\left[1_M' \otimes I_K\right]' \\
&= \left[1_M' \otimes I_K\right]\left\{\sum_{h=0}^{H-1}\Pi^h\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2 \otimes \Sigma\right]\left(\Pi^h\right)'\right. \\
&\quad - \sigma_{ii}^{-1}\left.\left\{\sum_{h=0}^{H-1}\Pi^h\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2 \otimes \left(\Sigma e_i\right)^2\right]\left(\Pi^h\right)'\right\}\right\}\left[1_M' \otimes I_K\right]'.
\end{aligned} \tag{2.47}
$$

Notice now, that the normality assumption allows for the computation of the conditional covariance matrix without any other knowledge about the conditioning variable than its distribution, thus enabling a univocal determination of the variance decomposition. Finally, define

$$
\begin{aligned}
\Delta_{t,\,i}\left(H\right) &= \text{cov}\left(\omega_t^{(y,\varepsilon)}\left(H\right)\right) - \text{cov}\left(\omega_t^{(y,\varepsilon,\,i)}\left(H\right)\right) \\
&= \left[1_M' \otimes I_K\right] \\
&\times \left\{\sigma_{ii}^{-1}\left\{\sum_{h=0}^{H-1}\Pi^h\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2 \otimes \left(\Sigma e_i\right)^2\right]\left(\Pi^h\right)'\right\}\right\} \\
&\times \left[1_M' \otimes I_K\right]'.
\end{aligned} \tag{2.48}
$$

The quantity above is always positive and shows that conditioning on the $i-th$ variable reduces the forecast error variance; thus, every $(j,\,j)$ element of $\Delta_{t,\,i}\left(H\right)$ is the share of variance in the $j-th$ equation for which the $i-th$ variable is responsible. Scaling this quantity with the total variance of the forecast error of the $j-th$ equation returns the GVD that constitutes the building block of the Diebold Yilmaz measures of connectedness, that is, the share of variance from variable i to variable j:

$$d_{ij,\,t}\left(H\right) = \frac{e_j'\left[1_M'\otimes I_K\right]\left\{\sum_{h=0}^{H-1}\Pi^h\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2\otimes\left(\Sigma e_i\right)^2\right]\left(\Pi^h\right)'\right\}\left[1_M'\otimes I_K\right]' e_j}{\sigma_{ii}e_j'\left[1_M'\otimes I_K\right]\left\{\sum_{h=0}^{H-1}\Pi^h\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2\otimes\Sigma\right]\left(\Pi^h\right)'\right\}\left[1_M'\otimes I_K\right]' e_j}. \tag{2.49}$$

Since the generalized variance decompositions do not sum to one, one of the following normalizations is usually employed to transform connected measures into indexes:

$$\tilde{d}_{ij,\,t}\left(H\right) = \frac{d_{ij,\,t}\left(H\right)}{\sum_{i=1}^{K} d_{ij,\,t}\left(H\right)}, \tag{2.50}$$

2.4 Generalizations

In this section, the results obtained for the MSA(M)-VAR(1) are extended in order to accommodate for a wide range of MS-VAR specifications. It will also be shown that the measure will remain invariant to the removal of restrictions on both the mean and the intercept.

2.4.1 Switching Intercept: MSIA(M)-VAR(1)

Consider the model examined in the previous section, but this time allow for an intercept term that is conditional on the state variable:

$$\begin{aligned} y_t &= \nu\left(s_t\right) + A\left(s_t\right)y_{t-1} + \varepsilon_t \\ \xi_t &= F\xi_{t-1} + v_t, \end{aligned} \tag{2.51}$$

Employing the Karlsen state space representation leads to:

$$
\begin{aligned}
\psi_t &= \xi_t \otimes y_t \\
&= [F\xi_{t-1} + v_t] \otimes [\nu(s_t) + A(s_t) y_{t-1} + \varepsilon_t] \\
&= (F\xi_{t-1}) \otimes \nu(s_t) + v_t \otimes \nu(s_t) + (F\xi_{t-1}) \otimes (A(s_t) y_{t-1}) \\
&+ v_t \otimes (A(s_t) y_{t-1}) + (F\xi_{t-1}) \otimes \varepsilon_t + v_t \otimes \varepsilon_t \\
&= (F\xi_{t-1}) \otimes \nu(s_t) + (F \otimes A(s_t))(\xi_{t-1} \otimes y_{t-1}) \\
&+ v_t \otimes \nu(s_t) + v_t \otimes (A(s_t) y_{t-1}) + (F\xi_{t-1} + v_t) \otimes \varepsilon_t \\
&= (F \otimes \nu(s_t))\, \xi_{t-1} + (F \otimes A(s_t))(\xi_{t-1} \otimes y_{t-1}) \\
&+ v_t \otimes \nu(s_t) + v_t \otimes (A(s_t) y_{t-1}) + (F\xi_{t-1} + v_t) \otimes \varepsilon_t \\
&= G\xi_{t-1} + \Pi\psi_{t-1} + \epsilon_t, \qquad (2.52)
\end{aligned}
$$

where an application of the results in the Appendix 2.C lets G to be expressed as

$$
G = F \otimes \nu(s_t) = \begin{bmatrix} p_{11}\nu_1 & p_{21}\nu_1 & \dots & p_{M1}\nu_1 \\ p_{12}\nu_2 & p_{22}\nu_2 & \dots & p_{M2}\nu_2 \\ \vdots & \vdots & \ddots & \vdots \\ p_{1M}\nu_M & p_{2M}\nu_M & \dots & p_{MM}\nu_M \end{bmatrix}, \qquad (2.53)
$$

Π can be still written as

$$
\Pi = F \otimes A(s_t) = \begin{bmatrix} p_{11}A_1 & p_{21}A_1 & \dots & p_{M1}A_1 \\ p_{12}A_2 & p_{22}A_2 & \dots & p_{M2}A_2 \\ \vdots & \vdots & \ddots & \vdots \\ p_{1M}A_M & p_{2M}A_M & \dots & p_{MM}A_M \end{bmatrix} \qquad (2.54)
$$

and

$$
\epsilon_t = v_t \otimes \nu(s_t) + v_t \otimes (A(s_t) y_{t-1}) + (F\xi_{t-1} + v_t) \otimes \varepsilon_t \qquad (2.55)
$$

is a martingale difference sequence.

To make the specification tractable and get rid of the intercept term, a new state space representation will be employed as in

$$\begin{aligned}\psi_t^* = \begin{bmatrix} \psi_t \\ \xi_t \end{bmatrix} &= \begin{bmatrix} \Pi & G \\ 0 & F \end{bmatrix} \begin{bmatrix} \psi_{t-1} \\ \xi_{t-1} \end{bmatrix} + \begin{bmatrix} \epsilon_t \\ v_t \end{bmatrix} \\ &= \Pi^* \psi_{t-1}^* + \epsilon_t^*, \end{aligned} \tag{2.56}$$

so that the moving average representation of ψ can be written in the following form:

$$\psi_{t+H}^* = \Pi^{*H} \psi_t^* + \sum_{h=0}^{H-1} \Pi^{*h} \epsilon_{t+H-h}^*. \tag{2.57}$$

The forecast of ψ^* and the forecast error of the linear state space representation can thus be expressed as

$$E_t \left[\psi_{t+H}^* \right] = \Pi^{*H} \psi_t^* \tag{2.58}$$

and

$$\begin{aligned}\omega^{*(\psi)}(H) &= \psi_{t+H}^* - E_t \left[\psi_{t+H}^* \right] \\ &= \sum_{h=0}^{H-1} \Pi^{*h} \epsilon_{t+H-h}^* \\ &= \sum_{h=0}^{H-1} \Pi^{*h} \begin{bmatrix} v_t \otimes \nu(s_t) + v_t \otimes (A(s_t) y_{t-1}) + (F\xi_{t-1} + v_t) \otimes \varepsilon_t \\ v_t \end{bmatrix}, \end{aligned} \tag{2.59}$$

so that the moving average representation, the forecast and the forecast error of y can be written as

$$
\begin{aligned}
y_{t+H} &= \sum_{m=1}^{M} \xi_{m,\,t+H} y_{t+H} \\
&= \left[1'_M \otimes I_K \,:\, 0_{K,\,M}\right] \psi^*_{t+H} \\
&= \left[1'_M \otimes I_K \,:\, 0_{K,\,M}\right] \left[\Pi^{*H} \psi^*_t + \sum_{h=0}^{H-1} \Pi^{*h} \epsilon^*_{t+H-h}\right], \qquad (2.60)
\end{aligned}
$$

$$
\begin{aligned}
E_t\left[y_{t+H}\right] &= \sum_{m=1}^{M} E_t\left[\xi_{m,\,t+H} y_{t+H}\right] \\
&= \left[1'_M \otimes I_K \,:\, 0_{K,\,M}\right] E_t\left[\psi^*_{t+H}\right] \\
&= \left[1'_M \otimes I_K \,:\, 0_{K,\,M}\right] \Pi^{*H} \psi^*_t, \qquad (2.61)
\end{aligned}
$$

and

$$
\begin{aligned}
\omega^{*(y)}_t(H) &= y_{t+H} - E_t\left[y_{t+H}\right] \\
&= \left[1'_M \otimes I_K \,:\, 0_{K,\,M}\right] \omega^{*(\psi)}_t(H) \\
&= \left[1'_M \otimes I_K \,:\, 0_{K,\,M}\right] \left\{ \sum_{h=0}^{H-1} \Pi^{*h} \right. \\
&\quad \times \left. \left[\begin{array}{c} v_{t+H-h} \otimes \nu\left(s_{t+H-h}\right) + v_t \otimes \left(A\left(s_{t+H-h}\right) y_{t-H-h}\right) \\ + \left(F\xi_{t+H-h-1} + v_{t+H-h}\right) \otimes \varepsilon_{t+H-h} \\ \\ v_t \end{array} \right] \right\} \\
&= \left[1'_M \otimes I_K \,:\, 0_{K,\,M}\right] \left\{ \left\{ \sum_{h=0}^{H-1} \Pi^{*h} \left[\begin{array}{c} v_{t+H-h} \otimes \nu\left(s_{t+H-h}\right) \\ + v_{t+H-h} \otimes \left(A\left(s_t\right) y_{t+H-h-1}\right) \\ \\ v_t \end{array} \right] \right. \right.
\end{aligned}
$$

$$
\begin{aligned}
&+\sum_{h=0}^{H-1} \Pi^{*h} \left[\begin{array}{c} v_{t+H-h} \otimes \varepsilon_{t+H-h} \\ 0 \end{array} \right] \Bigg] \Bigg\} \\
&+\sum_{h=1}^{H} \Pi^{*h} \left[\begin{array}{c} \left(F\xi_{t+H-h-1}\right) \otimes \varepsilon_{t+H-h} \\ 0 \end{array} \right] \Bigg] \Bigg\} \\
&= \omega_t^{(y,\,\varepsilon,\,v)}\left(H\right) + \omega_t^{(y,\,\varepsilon)}\left(H\right).
\end{aligned} \tag{2.62}
$$

As before, the quantity of interest is $\omega_t^{*(y,\,\varepsilon)}\left(H\right)$, that is, the normal component of the forecast error, although this time

$$
\begin{aligned}
\omega_t^{*(y,\,\varepsilon)}\left(H\right) &= \left[1_M' \otimes I_K \;:\; 0_{K,\,M}\right] \sum_{h=1}^{H} \Pi^{*h} \left[\begin{array}{c} \left(F\xi_{t+H-h-1}\right) \otimes \varepsilon_{t+H-h} \\ 0 \end{array} \right] \\
&= \left[1_M' \otimes I_K \;:\; 0_{K,\,M}\right] \sum_{h=1}^{H} \left[\begin{array}{cc} \Pi & G \\ 0 & F \end{array} \right]^h \left[\begin{array}{c} \left(F\xi_{t+H-h-1}\right) \otimes \varepsilon_{t+H-h} \\ 0 \end{array} \right].
\end{aligned} \tag{2.63}
$$

Then, since the lower left quadrant of Π^{*h} is a matrix of zeros and the vector post multiplying it has the last Q terms equal to zero,

$$
\begin{aligned}
\omega_t^{*(y,\,\varepsilon)}\left(H\right) &= \left[1_M' \otimes I_K \;:\; 0_{K,\,M}\right] \sum_{h=1}^{H} \Pi^{*h} \left[\begin{array}{c} \left(F\xi_{t+H-h-1}\right) \otimes \varepsilon_{t+H-h} \\ 0 \end{array} \right] \\
&= \left[1_M' \otimes I_K\right] \sum_{h=1}^{H} \Pi^{h} \left[\left(F\xi_{t+H-h-1}\right) \otimes \varepsilon_{t+H-h}\right] \\
&= \omega_t^{(y,\,\varepsilon)}\left(H\right),
\end{aligned} \tag{2.64}
$$

which is equal to the forecast error of the MSA specification and, as such, an application of the theory explained for the MSA case would lead to

$$d_{ij,\,t}\left(H\right)=\frac{e_j'\left[1_M'\otimes I_K\right]\left\{\sum_{h=0}^{H-1}\Pi^h\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2\otimes\left(\Sigma e_i\right)^2\right]\left(\Pi^h\right)'\right\}\left[1_M'\otimes I_K\right]'e_j}{\sigma_{ii}e_j'\left[1_M'\otimes I_K\right]\left\{\sum_{h=0}^{H-1}\Pi^h\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2\otimes\Sigma\right]\left(\Pi^h\right)'\right\}\left[1_M'\otimes I_K\right]'e_j}. \tag{2.65}$$

2.4.2 MSMA(K)-VAR(1) Models

For MSMA(K)-VAR(1) models, the procedures are quite similar to the MSA case, but before employing them, the state space needs to be extended to take into account the $p+1$ periods on which the Markov chain is conditioned. For $p=1$, the original Markov chain is of order 2 and the state space can be extended with the formulae of section 2, that is:

$$\xi_t=\xi_t^{(1)}\otimes\xi_{t-1}^{(1)}, \tag{2.66}$$

where $\xi_t^{(1)}\in\mathcal{M}\left(M,\,1\right)$, $\xi_t\in\mathcal{M}\left(N,\,1\right)$ and $N=M^{p+1}$. Then

$$F=\left[\mathrm{diag}\left(\mathrm{vec}P\otimes 1_{M^{p-1}}\right)\right]\left(1_M\otimes I_{M^p}\otimes 1_M'\right). \tag{2.67}$$

Following this adaptation, the state space can now be represented as

$$\begin{aligned} y_t-\mu_0\left(s_t\right)&=A\left(s_t\right)\left(y_{t-1}-\mu_1\left(s_t\right)\right)+\varepsilon_t\\ \xi_t&=F\xi_{t-1}+v_t, \end{aligned} \tag{2.68}$$

and can be rewritten into

$$
\begin{aligned}
y_t &= \mu_0\left(s_t\right) + z_t \\
z_t &= A\left(s_t\right) z_{t-1} + \varepsilon_t \\
\xi_t &= F\xi_{t-1} + v_t.
\end{aligned} \tag{2.69}
$$

In this way, the state space representation becomes

$$
\begin{aligned}
\psi_t &= \xi_t \otimes y_t \\
&= \xi_t \otimes \mu_0\left(s_t\right) + \xi_t \otimes z_t \\
&= \left[F\xi_{t-1} + v_t\right] \otimes \left[\mu_0\left(s_t\right) + A\left(s_t\right) z_{t-1} + \varepsilon_t\right] \\
&= \left(F\xi_{t-1}\right) \otimes \mu_0\left(s_t\right) + v_t \otimes \mu_0\left(s_t\right) + \left(F\xi_{t-1}\right) \otimes \left(A\left(s_t\right) z_{t-1}\right) \\
&+ v_t \otimes \left(A\left(s_t\right) z_{t-1}\right) + \left(F\xi_{t-1}\right) \otimes \varepsilon_t + v_t \otimes \varepsilon_t \\
&= \left(F\xi_{t-1}\right) \otimes \mu_0\left(s_t\right) + \left(F \otimes A\left(s_t\right)\right)\left(\xi_{t-1} \otimes z_{t-1}\right) \\
&+ v_t \otimes \mu_0\left(s_t\right) + v_t \otimes \left(A\left(s_t\right) z_{t-1}\right) + \left(F\xi_{t-1} + v_t\right) \otimes \varepsilon_t \\
&= \left(F \otimes \mu_0\left(s_t\right)\right)\xi_{t-1} + \left(F \otimes A\left(s_t\right)\right)\left(\xi_{t-1} \otimes z_{t-1}\right) \\
&+ v_t \otimes \mu_0\left(s_t\right) + v_t \otimes \left(A\left(s_t\right) z_{t-1}\right) + \left(F\xi_{t-1} + v_t\right) \otimes \varepsilon_t \\
&= Q\xi_{t-1} + \Pi\psi_{t-1}^{*} + \gamma_t + \epsilon_t,
\end{aligned} \tag{2.70}
$$

where an application of the results in the Appendix 2.C lets Q to be expressed as

$$
Q = F \otimes \mu_0\left(s_t\right) = \begin{bmatrix}
p_{11}\mu_{0,1} & p_{21}\mu_{0,1} & \cdots & p_{M1}\mu_{0,1} \\
p_{12}\mu_{0,2} & p_{22}\mu_{0,2} & \cdots & p_{M2}\mu_{0,2} \\
\vdots & \vdots & \ddots & \vdots \\
p_{1M}\mu_{0,M} & p_{2M}\mu_{0,M} & \cdots & p_{MM}\mu_{0,M}
\end{bmatrix}, \tag{2.71}
$$

Π can be still written as

$$\Pi = F \otimes A\left(s_t\right) = \begin{bmatrix} p_{11}A_1 & p_{21}A_1 & \dots & p_{M1}A_1 \\ p_{12}A_2 & p_{22}A_2 & \dots & p_{M2}A_2 \\ \vdots & \vdots & \ddots & \vdots \\ p_{1M}A_M & p_{2M}A_M & \dots & p_{MM}A_M \end{bmatrix}, \tag{2.72}$$

$$\epsilon_t = v_t \otimes \mu_0\left(s_t\right) + v_t \otimes \left(A\left(s_t\right) z_{t-1}\right) + \left(F\xi_{t-1} + v_t\right) \otimes \varepsilon_t \tag{2.73}$$

is a martingale difference sequence, $\psi_t^* = \xi_t \otimes z_t$ and $\gamma_t = v_t \otimes \mu_0\left(s_t\right)$.

Considering now the last equality in the previous equation, we further develop the two terms as

$$\begin{aligned} \xi_{t+H} \otimes \mu_0\left(s_{t+H}\right) &= \left(F^H\xi_t + \sum_{h=0}^{H-1} F^h v_{t+H-h}\right) \otimes \mu_0\left(s_{t+H}\right) \\ &= \left(F^H\xi_t\right) \otimes \mu_0\left(s_{t+H}\right) + \left(\sum_{h=0}^{H-1} F^h v_{t+H-h}\right) \otimes \mu_0\left(s_{t+H}\right) \\ &= \left(FF^{H-1}\xi_t\right) \otimes \mu_0\left(s_{t+H}\right) + \left(\sum_{h=0}^{H-1} FF^{h-1} v_{t+H-h}\right) \otimes \mu_0\left(s_{t+H}\right) \\ &= QF^{H-1}\xi_t + \sum_{h=0}^{H-1} QF^{h-1} v_{t+H-h}, \end{aligned} \tag{2.74}$$

and

$$\begin{aligned} \psi_t^* &= \xi_t \otimes z_t \\ &= \left[F\xi_{t-1} + v_t\right] \otimes \left[A\left(s_t\right) z_{t-1} + \varepsilon_t\right] \\ &= \left(F\xi_{t-1}\right) \otimes \left(A\left(s_t\right) z_{t-1}\right) + v_t \otimes \left(A\left(s_t\right) z_{t-1}\right) + \left(F\xi_{t-1}\right) \otimes \varepsilon_t + v_t \otimes \varepsilon_t \\ &= \Pi\psi_{t-1}^* + \epsilon_t, \end{aligned} \tag{2.75}$$

would allow for

$$\psi_{t+H}^{*} = \Pi^{H}\psi_{t}^{*} + \sum_{h=0}^{H-1}\Pi^{h}\epsilon_{t+H-h}, \tag{2.76}$$

so that the moving average, the forecast and the forecast error of the state space representation could be written as

$$\psi_{t+H} = QF^{H-1}\xi_{t} + \sum_{h=0}^{H-1}QF^{h-1}v_{t+H-h} + \Pi^{H}\psi_{t}^{*} + \sum_{h=0}^{H-1}\Pi^{h}\epsilon_{t+H-h}, \tag{2.77}$$

$$E_{t}\left[\psi_{t+H}\right] = QF^{H-1}\xi_{t} + \Pi^{H}\psi_{t}^{*}, \tag{2.78}$$

$$\begin{aligned} \omega^{(\psi)}\left(H\right) &= \psi_{t+H} - E_{t}\left[\psi_{t+H}\right] \\ &= \sum_{h=0}^{H-1}QF^{h-1}v_{t+H-h} + \sum_{h=0}^{H-1}\Pi^{h}\epsilon_{t+H-h} \\ &= \sum_{h=0}^{H-1}QF^{h-1}v_{t+H-h} \\ &+ \sum_{h=0}^{H-1}\Pi^{h}\left[v_{t+H-h}\otimes\left(A\left(s_{t}\right)y_{t+H-h-1}\right)\right. \\ &+ \left.\left(F\xi_{t+H-h-1} + v_{t+H-h}\right)\otimes\varepsilon_{t+H-h}\right]. \end{aligned} \tag{2.79}$$

These considerations allow to write the moving average representation, the forecast and the forecast error of the original process as

$$\begin{aligned} y_{t+H} &= \sum_{k=1}^{K}\xi_{k,\,t+H}y_{t+H} \\ &= \left[1_{N}'\otimes I_{K}\right]\psi_{t+H} \end{aligned}$$

$$= [1'_N \otimes I_K] \left[QF^{H-1}\xi_t + \sum_{h=0}^{H-1} QF^{h-1}v_{t+H-h} + \Pi^H \psi_t + \sum_{h=0}^{H-1} \Pi^h \epsilon_{t+H-h} \right], \tag{2.80}$$

$$\begin{aligned} E_t[y_{t+H}] &= \sum_{k=1}^{K} E_t[\xi_{k,\,t+H} y_{t+H}] \\ &= [1'_N \otimes I_K] E_t[\psi_{t+H}] \\ &= [1'_N \otimes I_K] \left(QF^{H-1}\xi_t + \Pi^H \psi_t \right), \end{aligned} \tag{2.81}$$

while

$$\begin{aligned} \omega_t^{(y)}(H) &= y_{t+H} - E_t[y_{t+H}] \\ &= [1'_N \otimes I_K] \, \omega_t^{(\psi)}(H) \\ &= [1'_N \otimes I_K] \left\{ \sum_{h=0}^{H-1} QF^{h-1} v_{t+H-h} \right. \\ &+ \sum_{h=0}^{H-1} \Pi^h [v_{t+H-h} \otimes (A(s_{t+H-h}) y_{t+H-h-1}) \\ &+ \left. (F\xi_{t+H-h-1} + v_{t+h}) \otimes \varepsilon_{t+H-h}] \right\} \\ &= [1'_N \otimes I_K] \left\{ \left\{ \sum_{h=0}^{H-1} QF^{h-1} v_{t+H-h} \right. \right. \\ &+ \left. \sum_{h=0}^{H-1} \Pi^h [v_{t+h} \otimes (A(s_{t+H-h}) y_{t+H-h-1})] + \sum_{h=0}^{H-1} \Pi^h [v_{t+H-h} \otimes \varepsilon_{t+H-h}] \right\} \\ &+ \left. \sum_{h=1}^{H} \Pi^h [(F\xi_{t+H-h-1}) \otimes \varepsilon_{t+H-h}] \right\} \\ &= \omega_t^{(y,\,\varepsilon,\,v)}(H) + \omega_t^{(y,\,\varepsilon)}(H). \end{aligned} \tag{2.82}$$

Applying the methods developed for the MSA case on the normal component of the forecast error, the scaled variance decompositions can be computed as

$$d_{ij,\,t}\left(H\right)=\frac{e_{j}^{\prime}\left[1_{N}^{\prime}\otimes I_{K}\right]\left\{\sum_{h=0}^{H-1}\Pi^{h}\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)^{2}\otimes\left(\Sigma e_{i}\right)^{2}\right]\left(\Pi^{h}\right)^{\prime}\right\}\left[1_{N}^{\prime}\otimes I_{K}\right]^{\prime}e_{j}}{\sigma_{ii}e_{j}^{\prime}\left[1_{N}^{\prime}\otimes I_{K}\right]\left\{\sum_{h=0}^{H-1}\Pi^{h}\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)^{2}\otimes\Sigma\right]\left(\Pi^{h}\right)^{\prime}\right\}\left[1_{N}^{\prime}\otimes I_{K}\right]^{\prime}e_{j}}. \tag{2.83}$$

One should notice now that, although functionally similar, the formula for the variance decompositions of MSMA models differs from the formula for the MSA models in the number of states, because MSMA models require the extension of the state space.

2.4.3 MSAH

To extend the basic MSA(M)-VAR(1) framework in order to allow for Markov switching heteroscedasticity, consider the following specification:

$$\begin{aligned} y_{t} &= A\left(s_{t}\right)y_{t-1}+\Sigma^{\frac{1}{2}}\left(s_{t}\right)u_{t} \\ \xi_{t} &= F\xi_{t-1}+v_{t}, \end{aligned} \tag{2.84}$$

where $F=P^{\prime}$, u_{t} is a standard normal error term and v_{t} is a martingale difference sequence. Then the Karlsen linear state space representation can be written as

$$\begin{aligned} \psi_{t} &= \xi_{t}\otimes y_{t} \\ &= \left[F\xi_{t-1}+v_{t}\right]\otimes\left[A\left(s_{t}\right)y_{t-1}+\Sigma^{\frac{1}{2}}\left(s_{t}\right)u_{t}\right] \\ &= \left(F\xi_{t-1}\right)\otimes\left(A\left(s_{t}\right)y_{t-1}\right)+v_{t}\otimes\left(A\left(s_{t}\right)y_{t-1}\right) \\ &\quad+\left(F\xi_{t-1}\right)\otimes\left(\Sigma^{\frac{1}{2}}\left(s_{t}\right)u_{t}\right)+v_{t}\otimes\left(\Sigma^{\frac{1}{2}}\left(s_{t}\right)u_{t}\right) \\ &= \left(F\otimes A\left(s_{t}\right)\right)\left(\xi_{t-1}\otimes y_{t-1}\right)+v_{t}\otimes\left(A\left(s_{t}\right)y_{t-1}\right) \end{aligned}$$

$$+\left(F\otimes\Sigma^{\frac{1}{2}}(s_t)\right)(\xi_{t-1}\otimes u_t)+v_t\otimes\left(\Sigma^{\frac{1}{2}}(s_t)\,u_t\right)$$
$$=\Pi\psi_{t-1}+\epsilon_t, \tag{2.85}$$

where

$$\epsilon_{t,\,s_t}=v_t\otimes(A(s_t)\,y_{t-1})+\left(F\otimes\Sigma^{\frac{1}{2}}(s_t)\right)(\xi_{t-1}\otimes u_t)+v_t\otimes\left(\Sigma^{\frac{1}{2}}(s_t)\,u_t\right) \tag{2.86}$$

and, as usual,

$$\Pi=F\otimes A(s_t)=\begin{bmatrix} p_{11}A_1 & \dots & p_{M1}A_1 \\ \vdots & \ddots & \vdots \\ p_{1M}A_M & \dots & p_{MM}A_M \end{bmatrix}. \tag{2.87}$$

The forecast of the state space can then be written as

$$\psi_{t+H}=\Pi^H\psi_t+\sum_{h=0}^{H-1}\Pi^h\epsilon_{t+H-h,\,s_{t+H-h}}, \tag{2.88}$$

with a forecast error

$$\begin{aligned}\omega^{(\psi)}(H)&=\psi_{t+H}-E_t[\psi_{t+H}]\\&=\sum_{h=0}^{H-1}\Pi^h\epsilon_{t+H-h,\,s_{t+H-h}}\\&=\sum_{h=0}^{H-1}\Pi^h\left[v_{t+H-h}\otimes(A(s_{t+H-h})\,y_{t+H-h-1})\right.\\&\quad\left.+(F\xi_{t+H-h-1}+v_{t+H-h})\otimes\left(\Sigma^{\frac{1}{2}}(s_{t+H-h})\,u_{t+H-h}\right)\right].\end{aligned} \tag{2.89}$$

The moving average and the forecast of y can be expressed as

$$
\begin{aligned}
y_{t+H} &= \sum_{k=1}^{K} \xi_{k,\,t+H} y_{t+H} \\
&= \left[1_K' \otimes I_M\right] \psi_{t+H} \\
&= \left[1_K' \otimes I_M\right] \left[\Pi^H \psi_t + \sum_{h=0}^{H-1} \Pi^h \epsilon_{t+H-h,\, s_{t+H-h}}\right],
\end{aligned} \tag{2.90}
$$

$$
\begin{aligned}
E_t\left[y_{t+H}\right] &= \sum_{k=1}^{K} E_t\left[\xi_{k,\,t+H} y_{t+H}\right] \\
&= \left[1_K' \otimes I_M\right] E_t\left[\psi_{t+H}\right] \\
&= \left[1_K' \otimes I_M\right] \Pi^H \psi_t,
\end{aligned} \tag{2.91}
$$

so that the total forecast error is

$$
\begin{aligned}
\omega_t^{(y)}(H) &= y_{t+H} - E_t\left[y_{t+H}\right] \\
&= \left[1_M' \otimes I_K\right] \omega_t^{(\psi)}(H) \\
&= \left[1_M' \otimes I_K\right] \left\{ \sum_{h=0}^{H-1} \Pi^h \left[v_{t+H-h} \otimes \left(A\left(s_{t+H-h}\right) y_{t+H-h-1}\right)\right.\right. \\
&+ \left.\left. \left(F\xi_{t+H-h-1} + v_{t+H-h}\right) \otimes \left(\Sigma^{\frac{1}{2}}\left(s_t\right) u_{t+H-h}\right)\right]\right\} \\
&= \left[1_M' \otimes I_K\right] \left\{ \left\{ \sum_{h=0}^{H-1} \Pi^h \left[v_{t+H-h} \otimes \left(A\left(s_{t+H-h}\right) y_{t+H-h-1}\right)\right]\right.\right. \\
&+ \left. \sum_{h=0}^{H-1} \Pi^h \left[v_{t+H-h} \otimes \left(\Sigma^{\frac{1}{2}}\left(s_{t+H-h}\right) u_{t+H-h}\right)\right]\right\} \\
&+ \left. \sum_{h=1}^{H} \Pi^h \left[\left(F\xi_{t+H-h-1}\right) \otimes \left(\Sigma^{\frac{1}{2}}\left(s_{t+H-h}\right) u_{t+H-h}\right)\right]\right\} \\
&= \omega_t^{(y,\,\varepsilon,\,v)}(H) + \omega_t^{(y,\,u)}(H).
\end{aligned} \tag{2.92}
$$

It is now possible to compute the covariance of the Gaussian component of the forecast error as

$$
\begin{aligned}
\text{cov}\left(\omega_t^{(y,u)}(H)\right) &= E_t\left[\omega_t^{(y,u)}(H)\left(\omega_t^{(y,u)}(H)\right)'\right] \\
&= E_t\left\{[1_M' \otimes I_K]\right. \\
&\times \left\{\sum_{h=0}^{H-1}\Pi^h\left[\left(F\hat{\xi}_{t+H-h-1}\right)\otimes\left(\Sigma^{\frac{1}{2}}(s_{t+H-h})\,u_{t+H-h}\right)\right]\right. \\
&\times \left.\left[\left(F\hat{\xi}_{t+H-h-1}\right)\otimes\left(\Sigma^{\frac{1}{2}}(s_{t+H-h})\,u_{t+H-h}\right)\right]'(\Pi^h)'\right\} \\
&\times \left.[1_M' \otimes I_K]'\right\} \\
&= [1_M' \otimes I_K] \\
&\times E_t\left\{\sum_{h=0}^{H-1}\Pi^h\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\otimes\left(\Sigma^{\frac{1}{2}}(s_{t+H-h})\,u_{t+H-h}\right)\right]\right. \\
&\times \left.\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\otimes\left(\Sigma^{\frac{1}{2}}(s_{t+H-h})\,u_{t+H-h}\right)\right]'(\Pi^h)'\right\} \\
&\times [1_M' \otimes I_K]' \\
&= [1_M' \otimes I_K] \\
&\times \left\{\sum_{h=1}^{H}\Pi^h E_t\left\{\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\otimes\left(\Sigma^{\frac{1}{2}}(s_{t+H-h})\,u_{t+H-h}\right)\right]\right.\right. \\
&\times \left.\left.\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\otimes\left(\Sigma^{\frac{1}{2}}(s_{t+H-h})\,u_{t+H-h}\right)\right]'\right\}(\Pi^h)'\right\} \\
&\times [1_M' \otimes I_K]' \\
&= [1_M' \otimes I_K] \\
&\times \left\{\sum_{h=0}^{H-1}\Pi^h E_t\left\{\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\left(F^{H-h}\hat{\xi}_{t|\,t}\right)'\right]\right.\right. \\
&\otimes \left.\left.\left[\left(\Sigma^{\frac{1}{2}}(s_{t+H-h})\,u_t\right)\left(\Sigma^{\frac{1}{2}}(s_{t+H-h})\,u_{t+H-h}\right)'\right]\right\}(\Pi^h)'\right\} \\
&\times [1_M' \otimes I_K]'
\end{aligned}
$$

$$
\begin{aligned}
&= [1'_M \otimes I_K] \\
&\times \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|\,t} \right)^2 \right. \right. \\
&\otimes \; E_t \left[\left(\Sigma^{\frac{1}{2}} \left(s_{t+H-h} \right) u_{t+H-h} u'_{t+H-h} \Sigma^{\frac{1}{2}} \left(s_{t+H-h} \right)' \right) \right] \Big\} \left(\Pi^h \right)' \Big\} \\
&\times [1'_M \otimes I_K]' \\
&= [1'_M \otimes I_K] \\
&\times \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|\,t} \right)^2 \otimes \Sigma \left(s_{t+H-h} \right) \right\} \left(\Pi^h \right)' \right\} \\
&\times [1'_M \otimes I_K]'
\end{aligned} \tag{2.93}
$$

Applying Pesaran and Shin [34] leads to

$$
\begin{aligned}
\omega_t^{(y,\,u,\,i)}(H) &= [1'_M \otimes I_K] \\
&\times \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \left[\varepsilon_{t+H-h} - E \left(\varepsilon_{t+H-h} \middle|\, \varepsilon_{i,\,t+H-h} \right) \right] \right\} \right\},
\end{aligned} \tag{2.94}
$$

which, assuming

$$
\varepsilon_{t+H-h} \sim \mathcal{N} \left(0,\, \Sigma \left(s_{t+H-h} \right) \right), \tag{2.95}
$$

so that

$$
\begin{aligned}
E \left(\varepsilon_{t+H-h} \middle|\, \varepsilon_{i,\,t+H-h} \right) &= \left[\sigma_{ii}^{-1} \left(s_{t+H-h} \right) \Sigma \left(s_{t+H-h} \right) e_i \right] \varepsilon_{i,\,t+H-h} \\
&= \left[\sigma_{ii}^{-1} \left(s_{t+H-h} \right) \Sigma \left(s_{t+H-h} \right) e_i \right] \sigma_{ii}^{\frac{1}{2}} \left(s_{t+H-h} \right) u_{i,\,t+H-h},
\end{aligned} \tag{2.96}
$$

can be rewritten as

$$
\begin{aligned}
\omega_t^{(y,u,i)}(H) &= \left[1_M' \otimes I_K\right] \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|t} \right) \right.\right. \\
&\otimes \left.\left. \left[\varepsilon_{t+H-h} - \left(\sigma_{ii}^{-1}(s_{t+H-h}) \Sigma(s_{t+H-h}) e_i \right) \varepsilon_{i,t+H-h} \right] \right\}\right\} \\
&= \left[1_M' \otimes I_K\right] \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|t} \right) \right.\right. \\
&\otimes \left[\Sigma^{\frac{1}{2}}(s_{t+H-h}) u_{t+H-h} \right. \\
&- \left.\left.\left. \left(\sigma_{ii}^{-1}(s_{t+H-h}) \Sigma(s_{t+H-h}) e_i \right) \sigma_{ii}^{\frac{1}{2}}(s_{t+H-h}) u_{i,t+H-h} \right] \right\}\right\}. \qquad (2.97)
\end{aligned}
$$

This, along with the fact that u_t is serially uncorrelated, allows to write the covariance of the Gaussian component of the forecast error conditional on variable i as

$$
\begin{aligned}
\text{cov}\left(\omega_t^{(y,\varepsilon,i)}(H)\right) &= E_t\left[\omega_t^{(y,\varepsilon,i)}(H)\left(\omega_t^{(y,\varepsilon,i)}(H)\right)'\right] \\
&= \left[1_M' \otimes I_K\right] E_t \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|t} \right) \otimes \left[\Sigma^{\frac{1}{2}}(s_{t+H-h}) u_{t+H-h} \right.\right.\right. \\
&- \left.\left. \left(\sigma_{ii}^{-1}(s_{t+H-h}) \Sigma(s_{t+H-h}) e_i \right) \sigma_{ii}^{\frac{1}{2}}(s_{t+H-h}) u_{i,t+H-h} \right] \right\} \\
&\times \left\{ \left(F^{H-h} \hat{\xi}_{t|t} \right) \otimes \left[\Sigma^{\frac{1}{2}}(s_{t+H-h}) u_{t+H-h} \right.\right. \\
&- \left.\left. \left(\sigma_{ii}^{-1}(s_{t+H-h}) \Sigma(s_{t+H-h}) e_i \right) \sigma_{ii}^{\frac{1}{2}}(s_{t+H-h}) u_{i,t+H-h} \right] \right\}' \\
&\times \left.\left(\Pi^h\right)' \right\} \left[1_M' \otimes I_K\right]' \\
&= \left[1_M' \otimes I_K\right] E_t \left\{ \sum_{h=0}^{H-1} \Pi^h \left\{ \left(F^{H-h} \hat{\xi}_{t|t} \right) \otimes \Sigma^{\frac{1}{2}}(s_{t+H-h}) u_{t+H-h} \right.\right. \\
&- \left(F^{H-h} \hat{\xi}_{t|t} \right) \\
&\otimes \left.\left[\left(\sigma_{ii}^{-1}(s_{t+H-h}) \Sigma(s_{t+H-h}) e_i \right) \sigma_{ii}^{\frac{1}{2}}(s_{t+H-h}) u_{i,t+H-h} \right] \right\} \\
&\times \left\{ \left(F^{H-h} \hat{\xi}_{t|t} \right) \otimes \Sigma^{\frac{1}{2}}(s_{t+H-h}) u_{t+H-h} \right. \\
&- \left(F^{H-h} \hat{\xi}_{t|t} \right)
\end{aligned}
$$

$$
\begin{aligned}
&\otimes \left[\left(\sigma_{ii}^{-1}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\right)\sigma_{ii}^{\frac{1}{2}}\left(s_{t+H-h}\right)u_{i,\,t+H-h}\right]\Big\}'\\
&\times \left(\Pi^h\right)'\Big\}\left[1_M'\otimes I_K\right]'\\
&= \left[1_M'\otimes I_K\right]\left\{\sum_{h=0}^{H-1}\Pi^h\left[a_h+b_h-2c_h\right]\left(\Pi^h\right)'\right\}\left[1_M'\otimes I_K\right]',
\end{aligned}
\tag{2.98}
$$

where

$$
\begin{aligned}
a_h &= E_t\left\{\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\otimes\Sigma^{\frac{1}{2}}\left(s_{t+H-h}\right)u_{t+H-h}\right]\right.\\
&\times \left.\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\otimes\Sigma^{\frac{1}{2}}\left(s_{t+H-h}\right)u_{t+H-h}\right]'\right\}\\
&= \left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\left(F^{H-h}\hat{\xi}_{t|\,t}\right)'\right]\\
&\otimes E_t\left[\Sigma^{\frac{1}{2}}\left(s_{t+H-h}\right)u_{t+H-h}u_{t+H-h}'\left(\Sigma^{\frac{1}{2}}\left(s_{t+H-h}\right)\right)'\right]\\
&= \left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2\otimes\Sigma\left(s_{t+H-h}\right),
\end{aligned}
\tag{2.99}
$$

$$
\begin{aligned}
b_h &= E_t\left\{\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\otimes\left(\sigma_{ii}^{-1}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\right)\sigma_{ii}^{\frac{1}{2}}\left(s_{t+H-h}\right)u_{i,\,t+H-h}\right]\right.\\
&\times \left.\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\otimes\left(\sigma_{ii}^{-1}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\right)\sigma_{ii}^{\frac{1}{2}}\left(s_{t+H-h}\right)u_{i,\,t+H-h}\right]'\right\}\\
&= \left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\left(F^{H-h}\hat{\xi}_{t|\,t}\right)'\right]\otimes\Big\{\left(\sigma_{ii}^{-1}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\right)\sigma_{ii}\left(s_{t+H-h}\right)\\
&\times E_t\left[u_{i,\,t+H-h}u_{i,\,t+H-h}'\right]\left(\sigma_{ii}^{-1}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\right)'\Big\}\\
&= \left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2\\
&\otimes\left(\sigma_{ii}^{-1}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\sigma_{ii}\left(s_{t+H-h}\right)e_i'\left(\Sigma\left(s_{t+H-h}\right)\right)'\sigma_{ii}^{-1}\left(s_{t+H-h}\right)\right)\\
&= \left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2\otimes\left(\sigma_{ii}^{-1}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_ie_i'\left(\Sigma\left(s_{t+H-h}\right)\right)'\right)\\
&= \left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2\otimes\left(\sigma_{ii}^{-\frac{1}{2}}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\right)^2,
\end{aligned}
\tag{2.100}
$$

and

$$
\begin{aligned}
c_h &= E_t \left\{ \left[\left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \Sigma^{\frac{1}{2}} \left(s_{t+H-h} \right) u_{t+H-h} \right] \right. \\
&\times \left. \left[\left(F^{H-h} \hat{\xi}_{t|\,t} \right) \otimes \left(\sigma_{ii}^{-1} \left(s_{t+H-h} \right) \Sigma \left(s_{t+H-h} \right) e_i \right) \sigma_{ii}^{\frac{1}{2}} \left(s_{t+H-h} \right) u_{i,\,t+H-h} \right]' \right\} \\
&= \left[\left(F^{H-h} \hat{\xi}_{t|\,t} \right) \left(F^{H-h} \hat{\xi}_{t|\,t} \right)' \right] \\
&\otimes E_t \left\{ \Sigma^{\frac{1}{2}} \left(s_{t+H-h} \right) u_{t+H-h} \right. \\
&\times \left. \left[\left(\sigma_{ii}^{-1} \left(s_{t+H-h} \right) \Sigma \left(s_{t+H-h} \right) e_i \right) \sigma_{ii}^{\frac{1}{2}} \left(s_{t+H-h} \right) u_{i,\,t+H-h} \right]' \right\} \\
&= \left(F^{H-h} \hat{\xi}_{t|\,t} \right)^2 \\
&\otimes E_t \left\{ \Sigma^{\frac{1}{2}} \left(s_{t+H-h} \right) u_{t+H-h} u'_{i,\,t+H-h} \sigma_{ii}^{\frac{1}{2}} \left(s_{t+H-h} \right) \sigma_{ii}^{-1} \left(s_{t+H-h} \right) e'_i \left(\Sigma \left(s_{t+H-h} \right) \right)' \right\} \\
&= \left(F^{H-h} \hat{\xi}_{t|\,t} \right)^2 \\
&\otimes E_t \left\{ \Sigma^{\frac{1}{2}} \left(s_{t+H-h} \right) u_{t+H-h} u'_{t+H-h} \right. \\
&\times \left. \left(\Sigma^{\frac{1}{2}} \left(s_{t+H-h} \right) \right)' e_i \sigma_{ii}^{-1} \left(s_{t+H-h} \right) e'_i \left(\Sigma \left(s_{t+H-h} \right) \right)' \right\} \\
&= \left(F^{H-h} \hat{\xi}_{t|\,t} \right)^2 \otimes \left(\sigma_{ii}^{-1} \left(s_{t+H-h} \right) \Sigma \left(s_{t+H-h} \right) e_i e'_i \left(\Sigma \left(s_{t+H-h} \right)' \right) \right) \\
&= \left(F^{H-h} \hat{\xi}_{t|\,t} \right)^2 \otimes \left(\sigma_{ii}^{-\frac{1}{2}} \left(s_{t+H-h} \right) \Sigma \left(s_{t+H-h} \right) e_i \right)^2 , \qquad (2.101)
\end{aligned}
$$

so that

$$
\begin{aligned}
\operatorname{cov} \left(\omega_t^{(\varepsilon,\,i)} (H) \right) &= \left[1'_M \otimes I_K \right] \left\{ \sum_{h=0}^{H-1} \Pi^h \left[a_h + b_h - 2 c_h \right] \left(\Pi^h \right)' \right\} \left[1'_M \otimes I_K \right]' \\
&= \left[1'_M \otimes I_K \right] \left\{ \sum_{h=0}^{H-1} \Pi^h \left[\left(F^{H-h} \hat{\xi}_{t|\,t} \right)^2 \otimes \Sigma \left(s_{t+H-h} \right) \right] \left(\Pi^h \right)' \right. \\
&- \left\{ \sum_{h=0}^{H-1} \Pi^h \left[\left(F^{H-h} \hat{\xi}_{t|\,t} \right)^2 \right. \right. \\
&\left. \left. \otimes \left(\sigma_{ii}^{-\frac{1}{2}} \left(s_{t+H-h} \right) \Sigma \left(s_{t+H-h} \right) e_i \right)^2 \right] \left(\Pi^h \right)' \right\} \right\} \left[1'_M \otimes I_K \right]' . \qquad (2.102)
\end{aligned}
$$

Then, the notions described in the previous sections allow to write

$$d_{ij,\,t}\left(H\right) = \frac{e_j'\left[1_M' \otimes I_K\right]\left[\sum_{h=0}^{H-1}\Pi^h\Theta_{1,\,h}\left(\Pi^h\right)'\right]\left[1_M' \otimes I_K\right]' e_j}{e_j'\left[1_M' \otimes I_K\right]\left[\sum_{h=0}^{H-1}\Pi^h\Theta_{2,\,h}\left(\Pi^h\right)'\right]\left[1_M' \otimes I_K\right]' e_j}, \tag{2.103}$$

where

$$\Theta_{1,\,h} = \left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2 \otimes \left(\sigma_{ii}^{-\frac{1}{2}}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\right)^2 \tag{2.104}$$

and

$$\Theta_{2,\,h} = \left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2 \otimes \Sigma\left(s_{t+H-h}\right). \tag{2.105}$$

Consider now the following application of the distributive property of the transposition and the mixed product property under the Kronecker product:

$$\begin{aligned}\Theta_{1,\,h} &= \left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2 \otimes \left(\sigma_{ii}^{-\frac{1}{2}}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\right)^2 \\ &= \left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right) \otimes \left(\sigma_{ii}^{-\frac{1}{2}}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\right)\right] \\ &\times \left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right) \otimes \left(\sigma_{ii}^{-\frac{1}{2}}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\right)\right]'.\end{aligned} \tag{2.106}$$

The mixed product property also allows to write

$$\begin{aligned}\left(F^{H-h}\hat{\xi}_{t|\,t}\right) \otimes \left(\sigma_{ii}^{-\frac{1}{2}}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)e_i\right) = \Big\{&F^{H-h} \\ &\otimes \left[\sigma_{ii}^{-\frac{1}{2}}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)\right]\Big\} \\ &\times \left(\hat{\xi}_{t|\,t} \otimes e_i\right)\end{aligned} \tag{2.107}$$

and an application of Appendix 2.C allows for

$$
\begin{aligned}
\Lambda_i^{(H-h)} &= F^{H-h} \otimes \left[\sigma_{ii}^{-\frac{1}{2}}\left(s_{t+H-h}\right)\Sigma\left(s_{t+H-h}\right)\right] \\
&= \begin{bmatrix} F_{(1,\,1)}^{H-h}\sigma_{ii,\,1}^{-\frac{1}{2}}\Sigma_1 & \dots & F_{(M,\,1)}^{H-h}\sigma_{ii,\,1}^{-\frac{1}{2}}\Sigma_1 \\ \vdots & \ddots & \vdots \\ F_{(1,\,M)}^{h}\sigma_{ii,\,M}^{-\frac{1}{2}}\Sigma_M & \dots & F_{(M,\,M)}^{h}\sigma_{ii,\,M}^{-\frac{1}{2}}\Sigma_M \end{bmatrix},
\end{aligned}
\tag{2.108}
$$

so that

$$
\begin{aligned}
\Theta_{1,\,h} &= \Lambda^{(H-h)}\left(\hat{\xi}_{t|\,t}\otimes e_i\right)\left[\Lambda^{(H-h)}\left(\hat{\xi}_{t|\,t}\otimes e_i\right)\right]' \\
&= \left[\Lambda^{(H-h)}\left(\hat{\xi}_{t|\,t}\otimes e_i\right)\right]^2.
\end{aligned}
\tag{2.109}
$$

Similar considerations apply for

$$
\begin{aligned}
\Theta_{2,\,h} &= \left(F^{H-h}\hat{\xi}_{t|\,t}\right)^2 \otimes \Sigma\left(s_{t+H-h}\right) \\
&= \left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\left(F^{H-h}\hat{\xi}_{t|\,t}\right)'\right] \otimes \left[\Sigma\left(s_{t+H-h}\right) I_K\right] \\
&= \left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)\otimes \Sigma\left(s_{t+H-h}\right)\right]\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)'\otimes I_K\right] \\
&= \left[\Gamma^{(H-h)}\left(\hat{\xi}_{t|\,t}\otimes I_K\right)\right]\left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)'\otimes I_K\right],
\end{aligned}
\tag{2.110}
$$

where

$$
\Gamma^{(H-h)} = F^{H-h}\otimes\Sigma\left(s_{t+H-h}\right) = \begin{bmatrix} F_{(1,\,1)}^{H-h}\Sigma_1 & \dots & F_{(M,\,1)}^{H-h}\Sigma_1 \\ \vdots & \ddots & \vdots \\ F_{(1,\,M)}^{H-h}\Sigma_M & \dots & F_{(M,\,M)}^{H-h}\Sigma_M \end{bmatrix}.
\tag{2.111}
$$

2.4.4 MSA(K)-VAR(P)

Estimating MS-VAR models can be quite a demanding exercise, as the parameter space grows very fast when relaxing restrictions, especially in high dimensional systems. For this reason, in particular when using MSA-VAR specifications, a parsimonious approach should usually guide the practitioner in choosing vector autoregressions of order one. This notwithstanding, if the analyst decides to use higher autoregressive orders, she can simply adapt the results in the previous sections to the stacked VAR(1) model:

$$\bar{y}_t = \begin{bmatrix} y_t \\ y_{t-1} \\ \vdots \\ y_{t-P+1} \end{bmatrix} = \bar{\nu}(s_t) + \bar{A}(s_t)\,\bar{y}_{t-1} + \bar{\varepsilon}_t, \tag{2.112}$$

$$\bar{\nu}(s_t) = \begin{bmatrix} \nu(s_t) \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \tag{2.113}$$

$$\bar{A}(s_t) = \begin{bmatrix} A_1(s_t) & A_P(s_t) & \dots & A_P(s_t) \\ I_M & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & I_M & 0 \end{bmatrix}, \tag{2.114}$$

$$\bar{\varepsilon}_t = \begin{bmatrix} \varepsilon_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \tag{2.115}$$

2.5 A simulation-based practical application

To illustrate how the DY connectedness measures can behave under the MS-VAR assumption, we propose two simulation-based experiments.

The first example is based on Krolzig [27] estimation of a MSMH(3)-DVAR(1) model on international business cycle data. The second experiment is based on the more recent Guidolin et al. [18] (GOP) estimation of a MSIH(3)-VAR(1) model on a panel of corporate bond yields and Treasury yields. In both the experiments, the total connectedness measure will be computed assuming a forecast horizon (H) of ten periods and will employ smoothed states probabilities.

Simulating the Markov chain The first step to both Monte Carlo experiments is to use the transition matrix in each parametrization (see the parameters reported in the Appendices 2.D and 2.E) to simulate a path for the Markov chain. To do so, the following algorithm has been used:

1. select a starting point for the Markov chain ξ_0 (it makes sense to start with the most probable state);
2. detect the index i_{n-1} of the element equal to one of the previously extracted point of the Markov chain ξ_{n-1};
3. randomly extract r from the uniform distribution;

4. using the $i_{n-1}-th$ row of the transition matrix, compute a cumulative distribution;
5. choose, as i_n, the index for which the cumulative distribution is greater than r;
6. repeat from step 2 until reaching a desirable number of simulated points of the Markov chain.

Experiment with Krolzig parametrization In this experiment, the parameters obtained by Krolzig [27] in his MSMH(3)-DVAR(1) estimation of a multi-country growth model with Markov switching regimes will be used. In this study, Krolzig uses the data from OECD on real GNP of USA, Japan and West Germany and real GDP of the UK, Canada and Australia. These time series are then transformed in growth rates, before being employed in the analysis. The sample used by Krolzig spans from 1962:1 to 1991:4 and covers 120 quarters (excluding presample values).

The specification assumes three states of the world: state one, characterized by higher than average growth means, state two, characterized by average growth means, and state three, characterized by negative or lower than average means. DY connectedness is invariant with respect to the mean, thus, since the autoregressive parameters are constant in this specification, total connectedness reflects the dynamics of the state dependent variance covariance matrix.

Starting from Krolzig's parameters, reported in the Appendix 2.D, we will simulate the time series, from which inferred, predicted and smoothed state probabilities will be recovered. With these ingredients, a dynamic measure of total connectedness will be computed.

Using the simulated Markov chain is then possible to use the parameters in Krolzig [27] to simulate the time series using the formulae for each point of the MSMH(3)-VAR(1), so that, for each t:

$$y_t = \left(\sum_{m=1}^{M} \sum_{n(m)} \xi_{tn} X_n^{\mu} \right) \mu + \left(y'_{t-1} \otimes I_K \right) \alpha + u_t, \tag{2.116}$$

where

$$n(m) = (m-1)\frac{N}{M} + 1, \ldots, m\frac{N}{M}, \tag{2.117}$$

$$\xi_t = \xi_t^{(p+1)} = \otimes_{j=0}^{p} \xi_{t-j}^{(1)}, \tag{2.118}$$

$$\xi_t^{(1)} = \begin{bmatrix} I(s_t = 1) \\ \vdots \\ I(s_t = M) \end{bmatrix}, \tag{2.119}$$

$$X_n^{\mu} = (e'_n \otimes I_K) X^{\mu}, \tag{2.120}$$

$$X^{\mu} = -\sum_{l=0}^{p} L'_l \otimes A_l, \quad A_0 = -I_k, \tag{2.121}$$

$$L_l = 1'_{M^l} \otimes I_M \otimes 1_{M^{p-l}}, \quad l = 0, \ldots, p, \tag{2.122}$$

$$\mu = (\mu'_1, \ldots, \mu'_M)', \tag{2.123}$$

$$\alpha = \text{vec}(A), \tag{2.124}$$

and u_t is randomly extracted from $\mathcal{N}(0, \Omega_t)$, where

$$\Omega_t = \sum_{m=1}^{M} \sum_{n(m)} \xi_{tn} \Sigma_n. \tag{2.125}$$

The simulated time series can then be used to compute infered, predicted and smoothed state probabilities.

Total connectedness is then computed summing the non diagonal elements of the connectedness table computed using 2.126

$$d_{ij,\,t}\left(H\right) = \frac{e_j' \left[1_M' \otimes I_K\right] \left[\sum_{h=0}^{H-1} \Pi^h \theta_{1h,\,\text{Krolzig}}^2 \left(\Pi^h\right)'\right] \left[1_M' \otimes I_K\right]' e_j}{e_j' \left[1_M' \otimes I_K\right] \left\{\sum_{h=0}^{H-1} \Pi^h \theta_{2h,\,\text{Krolzig}} \left(\Pi^h\right)'\right\} \left[1_M' \otimes I_K\right]' e_j}, \tag{2.126}$$

where

$$\Pi = F \otimes A, \tag{2.127}$$

$$\theta_{1h,\,\text{Krolzig}} = \Lambda_i^{(H-h)} \left(\hat{\xi}_{t|\,t} \otimes e_i\right) \tag{2.128}$$

and

$$\theta_{2h,\,\text{Krolzig}} = \left[\Gamma^{(H-h)} \left(\hat{\xi}_{t|\,t} \otimes I_K\right)\right] \left[\left(F^{H-h} \hat{\xi}_{t|\,t}\right)' \otimes I_K\right]. \tag{2.129}$$

Figure 2.1 shows the plots of the elements of the Markov chain extracted from a 500 iterations simulation.

[Figure 2.1 about here.]

As expected, the most occurring state of the world is state two (normal growth), followed by respectively state one (higher than normal growth) and state three (negative or below average growth). As expected from the inspection of the transition matrix, the number of switches from another state (most likely from state two) to state three is higher than the switch from another state to state one, while, at the same time, the time spent in state one, once switched, is considerably longer than the time spent in state three. In fact, the simulated system visited state one 150 times, state two 275 times and state three 75 times, very much consistently both with the transition matrix and the ergodic state probabilities provided in Krolzig [27].

The plots for the simulated time series are shown in Figure 2.2, where it is possible to roughly observe the autoregressive pattern, along with small clusters of outlying points extracted from states one and three.

[Figure 2.2 about here.]

Figure 2.3 shows the plot of the total connectedness measure implied by the simulated series.

[Figure 2.3 about here.]

Figure 2.4 compares total connectedness with each item of the inferred state probabilities extracted from the simulated time series.

[Figure 2.4 about here.]

Interestingly, the lowest values for total connectedness happen during periods of higher than normal growth, while the highest values happen during recessions.

Experiment with Guidolin, Orlov and Pedio (GOP) parametrization The following experiment will use the parametrization obtained in Guidolin et al. [18]. They use data on corporate bonds yields and the Treasury yields curve from October, 8 2004 to December 28, 2012. With corporate bond yields observations, they build four portfolios with different ratings and maturities: (i) investment-grade short-term bonds (IGST), (ii) investment-grade long-term bonds (IGLT), (iii) non- investment-grade short-term bonds

(NIGST), and (iv) non-investment-grade long-term bonds (NIGLT). The treasury yields curve is then represented by 1-month, 1-year, 5- year, and 10-year weekly (Friday-to-Friday) constant maturity Treasury yields. GOP parameters for a MSIH(3)-VAR(1) are reported in the Appendix 2.E.

The specification assumes three states of the world: state one representing pre-crisis periods, state two representing crisis and state three representing post-crisis behaviour. States one and three are persistent states, with regime three being characterized by slightly lower yields, especially for the least risky bonds. State two is instead characterized by greater volatility and the phenomenon of flight to quality.

Using such parametrization the time series of the variables can be simulated using Krolzig [27]:

$$y_t = \sum_{m=1}^{M} \xi_{tm}\nu_m + \left(y'_{t-1} \otimes I_K\right)\alpha + u_t, \tag{2.130}$$

u_t being randomly extracted from $\mathcal{N}\left(0,\, \Omega_t\right)$, where

$$\Omega_t = \sum_{m=1}^{M} \xi_{tm}\Sigma_m. \tag{2.131}$$

Using the simulated time series, smoothed states probabilities are retrieved and total connectedness is computed as usual, that is, summing the non diagonal elements of the connectedness table computed using 2.132

$$d_{ij,\,t}\left(H\right) = \frac{e'_j \left[1'_M \otimes I_K\right] \left[\sum_{h=0}^{H-1} \Pi^h \theta^2_{1h,\,\mathrm{GOP}} \left(\Pi^h\right)'\right] \left[1'_M \otimes I_K\right]' e_j}{e'_j \left[1'_M \otimes I_K\right] \left[\sum_{h=0}^{H-1} \Pi^h \theta_{2h,\,\mathrm{GOP}} \left(\Pi^h\right)'\right] \left[1'_M \otimes I_K\right]' e_j}, \tag{2.132}$$

where

$$\Pi = F \otimes A, \tag{2.133}$$

$$\theta_{1h,\,\mathrm{GOP}} = \Lambda_i^{(H-h)} \left(\hat{\xi}_{t|\,t} \otimes e_i\right) \tag{2.134}$$

and

$$\theta_{2h,\,\mathrm{GOP}} = \left[\Gamma^{(H-h)} \left(\hat{\xi}_{t|\,t} \otimes I_K\right)\right] \left[\left(F^{H-h}\hat{\xi}_{t|\,t}\right)' \otimes I_K\right]. \tag{2.135}$$

Notice that, since connectedness is not affected either by means or intercepts, the formulae for computing the connectedness building blocks are the same for both the experiments.

Figure 2.5 shows the simulated path of the Markov chain for this experiment.

[Figure 2.5 about here.]

It is possible to observe that regimes one and three are indeed much more persistent than regime two. Moreover, regime two is most often (but not always) preceded by regime one and followed by regime three. The simulation visited regime one 252 times, regime two 49 times and regime three 199 times.

Figure 2.6 contains the plots of the simulated time series and Figure 2.7 shows the plot of the implied total connectedness.

[Figure 2.6 about here.]

[Figure 2.7 about here.]

Finally, Figure 2.8 compares total connectedness with smoothed probabilities filtered from the simulated series.

[Figure 2.8 about here.]

Again, total connectedness is much higher during the crisis regime, showing how this measure can capture in a single number useful features of the analyzed system. Interestingly enough, connectedness during pre-crisis is higher than connectedness in the post-crisis state, possibly indicating the success of the economic policies employed for crisis resolution.

2.6 Conclusions

Connectedness is a central topic in scientific inquiry, especially when large complex systems are involved, even though the formalization of such a concept can vary. The Diebold and Yilmaz measures try to capture connectedness using variance spillovers in a multivariate autoregressive setting. As the measures are functions of the system parameters, when the latter are made to vary, the measures mechanically display dynamics.

This paper addressed the problem of explicitly compute DY connectedness measures, when the source of parameter dynamics is a Markovian change of regimes, by deriving approximated closed formulae for the computation of generalized variance decompositions on which the connectedness measures are based. Such formulae are shown to be an aggregator of information about the distribution of a class of multivariate markov switching models.

As the framework depicted in Diebold and Yilmaz [12] suggests, further research can be certainly conducted, including, to cite a category, studies to come up with generalized variance decompositions for either different parameters dynamics and distributions of the error term.

Naturally, the methods described in this paper can be adapted to any form of empirical analysis that employs Markov switching vector autoregressions and can be useful, for example, to assess both systemic risk in portfolios of assets or spillovers occurring between assets and portfolios or any other system of variables.

As the scale of estimation increases, complexities arise, because the current tractable number of variables in the system and the number of restrictions on parameters fundamentally depend on, for example, things like sample size and computation power. As the scale of the model increases, more and more a priori choices (as divide and conquer decisions) have to be made, in order to perform parameter estimations.

As the simulation-based approach used in this paper attempted to show, connectedness measures can deliver meaningful and intuitive information about the observed system, which can be used by policymakers to assess and monitor the performance of their policies.

Bibliography

[1] Albert, J. H. and Chib, S. (1993). Bayes inference via gibbs sampling of autoregressive time series subject to markov mean and variance shifts. *Journal of Business & Economic Statistics*, 11(1):1–15.

[2] Ang, A. and Timmermann, A. (2012). Regime changes and financial markets. *Annu. Rev. Financ. Econ.*, 4(1):313–337.

[3] Baum, L. E., Petrie, T., Soules, G., and Weiss, N. (1970). A maximization technique occurring in the statistical analysis of probabilistic functions of markov chains. *The annals of mathematical statistics*, 41(1):164–171.

[4] Billio, M., Casarin, R., Ravazzolo, F., and Van Dijk, H. K. (2016). Interconnections between eurozone and us booms and busts using a bayesian panel markov-switching var model. *Journal of Applied Econometrics*, 31(7):1352–1370.

[5] Bostanci, G. and Yilmaz, K. (2015). How connected is the global sovereign credit risk network?

[6] Carter, C. K. and Kohn, R. (1994). On gibbs sampling for state space models. *Biometrika*, 81(3):541–553.

[7] Demirer, M., Diebold, F. X., Liu, L., and Yılmaz, K. (2017). Estimating global bank network connectedness. Technical report, National Bureau of Economic Research.

[8] Diebold, F. X., Liu, L., and Yilmaz, K. (2017). Commodity connectedness.

[9] Diebold, F. X. and Yilmaz, K. (2009). Measuring financial asset return and volatility spillovers, with application to global equity markets. *The Economic Journal*, 119(534):158–171.

[10] Diebold, F. X. and Yilmaz, K. (2012). Better to give than to receive: Predictive directional measurement of volatility spillovers. *International Journal of Forecasting*, 28(1):57–66.

[11] Diebold, F. X. and Yılmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134.

[12] Diebold, F. X. and Yilmaz, K. (2015a). *Financial and Macroeconomic Connectedness: A Network Approach to Measurement and Monitoring*. Oxford University Press, USA.

[13] Diebold, F. X. and Yilmaz, K. (2015b). Trans-atlantic equity volatility connectedness: Us and european financial institutions, 2004–2014. *Journal of Financial Econometrics*, 14(1):81–127.

[14] Ferrara, L., Marcellino, M., and Mogliani, M. (2015). Macroeconomic forecasting during the great recession: The return of non-linearity? *International Journal of Forecasting*, 31(3):664–679.

[15] Garcia, R. (1998). Asymptotic null distribution of the likelihood ratio test in markov switching models. *International Economic Review*, pages 763–788.

[16] Gaspar, R. M. (2012). Comment on better to give than to receive by francis x. diebold and kamil yilmaz. *International Journal of Forecasting*, 28(1):67–69.

[17] Granger, C. W. (2008). Non-linear models: Where do we go next-time varying parameter models? *Studies in Nonlinear Dynamics & Econometrics*, 12(3).

[18] Guidolin, M., Orlov, A. G., and Pedio, M. (2017). The impact of monetary policy on corporate bonds under regime shifts. *Journal of Banking & Finance*, 80:176–202.

[19] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica: Journal of the Econometric Society*, pages 357–384.

[20] Hamilton, J. D. (1994). *Time series analysis*. Princeton university press Princeton.

[21] Hamilton, J. D. (2010). Regime switching models. In *Macroeconometrics and time series analysis*, pages 202–209. Springer.

[22] Hansen, B. E. (1996). Inference when a nuisance parameter is not identified under the null hypothesis. *Econometrica: Journal of the econometric society*, pages 413–430.

[23] Karlsen, H. (1990). A class of non-linear time series models. *PhD diss., University of Bergen, Department of Mathematics*.

[24] Kim, C.-J. (1994). Dynamic linear models with markov-switching. *Journal of Econometrics*, 60(1-2):1–22.

[25] Kim, C.-J., Nelson, C. R., et al. (1999). State-space models with regime switching: classical and gibbs-sampling approaches with applications. *MIT Press Books*, 1.

[26] Koop, G., Pesaran, M. H., and Potter, S. M. (1996). Impulse response analysis in nonlinear multivariate models. *Journal of econometrics*, 74(1):119–147.

[27] Krolzig, H.-M. (2013). *Markov-switching vector autoregressions: Modelling, statistical inference, and application to business cycle analysis*, volume 454. Springer Science & Business Media.

[28] Krolzig, H.-M. et al. (2000). *Predicting Markov-switching vector autoregressive processes*. Nuffield College.

[29] Lindgren, G. (1978). Markov regime models for mixed distributions and switching regressions. *Scandinavian Journal of Statistics*, pages 81–91.

[30] Lütkepohl, H. (2005). *New introduction to multiple time series analysis*. Springer Science & Business Media.

[31] Magnus, J. R. (1988). *Linear Structures (Charles Griffin Series-Griffin's Statistical Monographs, No 42)*. Oxford University Press.

[32] McCulloch, R. E. and Tsay, R. S. (1994). Bayesian analysis of autoregressive time series via the gibbs sampler. *Journal of Time Series Analysis*, 15(2):235–250.

[33] Pesaran, B. and Pesaran, M. H. (2010). *Time series econometrics using Microfit 5.0: A user's manual*. Oxford University Press, Inc.

[34] Pesaran, H. H. and Shin, Y. (1998). Generalized impulse response analysis in linear multivariate models. *Economics letters*, 58(1):17–29.

[35] Sims, C. A. (1980). Macroeconomics and reality. *Econometrica: Journal of the Econometric Society*, pages 1–48.

[36] Sims, C. A., Waggoner, D. F., and Zha, T. (2008). Methods for inference in large multiple-equation markov-switching models. *Journal of Econometrics*, 146(2):255–274.

[37] Wold, H. O. (1960). A generalization of causal chain models (part iii of a triptych on causal chain systems). *Econometrica: Journal of the Econometric Society*, pages 443–463.

Appendix 2.A Proposition 1

Consider a primitive transition matrix P as defined in Krolzig, such as

$$P = \begin{bmatrix} p_{11} & p_{12} & \cdots & p_{1M} \\ p_{21} & p_{22} & \cdots & p_{2M} \\ \vdots & \vdots & \ddots & \vdots \\ p_{M1} & p_{M2} & \cdots & p_{MM} \end{bmatrix} : \sum_{j=1}^{M} p_{ij} = 1. \tag{A.1}$$

The Derivative Ttransition Matrix F should satisfy

$$\begin{aligned} Pr\big(\xi_{t+1} = \xi^{*}_{t+1} \big| \, \xi_t\big) = Pr & \left(\xi^{(1)}_{t+1} = \xi^{(1)^{*}}_{t+1} \big| \, \xi^{(1)}_{t}\right) \\ & \times Pr \left(\xi^{(1)}_{t} = \xi^{(1)^{*}}_{t}, \, \xi^{(1)}_{t-1} = \xi^{(1)^{*}}_{t-1}, \, \ldots, \, \xi^{(1)}_{t-p+1} = \xi^{(1)^{*}}_{t-p+1} \right. \\ & \left. \big| \, \xi^{(1)}_{t}, \, \xi^{(1)}_{t-1}, \, \ldots, \, \xi^{(1)}_{t-p}\right). \end{aligned} \tag{A.2}$$

To show it is true, consider each factor separately.

If $\xi^{(1)^{*}}_{t+1} = e_j$, given $\xi^{(1)}_{t} = e_i$, $Pr\left(\xi^{(1)}_{t+1} = e_j \big| \, \xi^{(1)}_{t} = e_i\right) = p_{ij}$. If $\xi^{(1)}_{t}$ is not specified as a $\xi^{(1)^{*}}_{t}$, then $Pr\left(\xi^{(1)}_{t+1} = e_j \big| \, \xi^{(1)}_{t}\right) = Pe_j$.

The second factor is a number that can be either 0 or 1 and is an element of $\xi^{(p)}_{t}$ determined by $\xi^{(1)^{*}}_{t}, \, \xi^{(1)^{*}}_{t-1}, \, \ldots, \, \xi^{(1)^{*}}_{t-p+1}$.

For example, for $M = 2$ and $p = 2$:

$$
\begin{array}{cccccc}
s_{t+1} & s_t & s_{t-1} & \text{First Factor} & \text{Second Factor} & Pr\left(\xi_{t+1} = \xi^*_{t+1} | \, \xi_t\right) \\
1 & 1 & 1 & p_{11} & \xi^{(2)*}_{1t} = \xi^{(1)}_{1t}\xi^{(1)}_{1t-1} & p_{11}\xi^{(1)}_{1t}\xi^{(1)}_{1t-1} \\
1 & 1 & 2 & p_{11} & \xi^{(2)*}_{2t} = \xi^{(1)}_{1t}\xi^{(1)}_{2t-1} & p_{11}\xi^{(1)}_{1t}\xi^{(1)}_{2t-1} \\
1 & 2 & 1 & p_{21} & \xi^{(2)*}_{3t} = \xi^{(1)}_{2t}\xi^{(1)}_{1t-1} & p_{21}\xi^{(1)}_{2t}\xi^{(1)}_{1t-1} \\
1 & 2 & 2 & p_{21} & \xi^{(2)*}_{4t} = \xi^{(1)}_{1t}\xi^{(1)}_{2t-1} & p_{21}\xi^{(1)}_{2t}\xi^{(1)}_{2t-1} \\
2 & 1 & 1 & p_{12} & \xi^{(2)*}_{1t} = \xi^{(1)}_{1t}\xi^{(1)}_{1t-1} & p_{12}\xi^{(1)}_{1t}\xi^{(1)}_{1t-1} \\
2 & 1 & 2 & p_{12} & \xi^{(2)*}_{2t} = \xi^{(1)}_{1t}\xi^{(1)}_{2t-1} & p_{12}\xi^{(1)}_{1t}\xi^{(1)}_{2t-1} \\
2 & 2 & 1 & p_{22} & \xi^{(2)*}_{3t} = \xi^{(1)}_{2t}\xi^{(1)}_{1t-1} & p_{22}\xi^{(1)}_{2t}\xi^{(1)}_{1t-1} \\
2 & 2 & 2 & p_{22} & \xi^{(2)*}_{4t} = \xi^{(1)}_{2t}\xi^{(1)}_{2t-1} & p_{22}\xi^{(1)}_{2t}\xi^{(1)}_{2t-1}
\end{array}
\tag{A.3}
$$

In this case,

$$
\begin{aligned}
diag\left(\text{vec}P \otimes I_{M^{p-1}}\right) &= diag\left(\begin{bmatrix} p_{11} \\ p_{21} \\ p_{12} \\ p_{22} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix}\right) \\
&= diag\left(\begin{bmatrix} p_{11} \\ p_{11} \\ p_{21} \\ p_{21} \\ p_{12} \\ p_{12} \\ p_{22} \\ p_{22} \end{bmatrix}\right)
\end{aligned}
$$

$$= \begin{bmatrix} p_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & p_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{21} & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & p_{21} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{12} & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & p_{12} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{22} & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{22} \end{bmatrix} \tag{A.4}$$

while

$$\begin{bmatrix} \xi_{1t}^{(1)}\xi_{1t-1}^{(1)} \\ \xi_{1t}^{(1)}\xi_{2t-1}^{(1)} \\ \xi_{2t}^{(1)}\xi_{1t-1}^{(1)} \\ \xi_{2t}^{(1)}\xi_{2t-1}^{(1)} \\ \xi_{1t}^{(1)}\xi_{1t-1}^{(1)} \\ \xi_{1t}^{(1)}\xi_{2t-1}^{(1)} \\ \xi_{2t}^{(1)}\xi_{1t-1}^{(1)} \\ \xi_{2t}^{(1)}\xi_{2t-1}^{(1)} \end{bmatrix} = 1_M \otimes \xi_t^{(p)} \tag{A.5}$$

and

$$1_M \otimes \xi_t^{(p)} = 1_M \otimes (I_{M^p} \otimes 1'_M)\, \xi_t$$

$$= 1_M \otimes (I_{M^p} \otimes 1'_M) \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \\ \xi_{3t} \\ \xi_{4t} \\ \xi_{5t} \\ \xi_{6t} \\ \xi_{7t} \\ \xi_{8t} \end{bmatrix}$$

$$= 1_M \otimes \left(\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix} \right) \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \\ \xi_{3t} \\ \xi_{4t} \\ \xi_{5t} \\ \xi_{6t} \\ \xi_{7t} \\ \xi_{8t} \end{bmatrix}$$

$$= 1_M \otimes \left(\begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \right) \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \\ \xi_{3t} \\ \xi_{4t} \\ \xi_{5t} \\ \xi_{6t} \\ \xi_{7t} \\ \xi_{8t} \end{bmatrix}$$

$$
\begin{aligned}
&= 1_M \otimes \begin{bmatrix} \xi_{1t} + \xi_{2t} \\ \xi_{3t} + \xi_{4t} \\ \xi_{5t} + \xi_{6t} \\ \xi_{7t} + \xi_{8t} \end{bmatrix} \\
&= 1_M \otimes \begin{bmatrix} \xi_{1t}^{(2)} \xi_{1t-2} + \xi_{1t}^{(2)} \xi_{2t-2} \\ \xi_{2t}^{(2)} \xi_{1t-2} + \xi_{2t}^{(2)} \xi_{2t-2} \\ \xi_{3t}^{(2)} \xi_{1t-2} + \xi_{1t}^{(2)} \xi_{2t-2} \\ \xi_{4t}^{(2)} \xi_{1t-2} + \xi_{4t}^{(2)} \xi_{2t-2} \end{bmatrix} \\
&= 1_M \otimes \begin{bmatrix} \xi_{1t}^{(2)} \\ \xi_{2t}^{(2)} \\ \xi_{3t}^{(2)} \\ \xi_{4t}^{(2)} \end{bmatrix} \\
&= 1_M \otimes \xi_t^{(2)} \qquad \text{(A.6)}
\end{aligned}
$$

Moreover,

$$
1_M \otimes I_{M^p} \otimes 1'_M = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix} \qquad \text{(A.7)}
$$

so that

$$
\begin{aligned}
F &= \mathrm{diag}\left(\mathrm{vec}P \otimes 1_{M^{p-1}}\right)\left(1_M \otimes I_{M^p} \otimes 1_M'\right) \\
&= \begin{bmatrix}
p_{11} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & p_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & p_{21} & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & p_{21} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_{12} & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & p_{12} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & p_{22} & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & p_{22}
\end{bmatrix} \\
&\times \begin{bmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1
\end{bmatrix} \\
&= \begin{bmatrix}
p_{11} & p_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & p_{11} & p_{11} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_{21} & p_{21} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & p_{21} & p_{21} \\
p_{12} & p_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & p_{12} & p_{12} & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & p_{22} & p_{22} & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & p_{22} & p_{22}
\end{bmatrix}
\end{aligned} \tag{A.8}
$$

and

$$
F\xi_t = \begin{bmatrix} p_{11} & p_{11} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{11} & p_{11} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{21} & p_{21} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{21} & p_{21} \\ p_{12} & p_{12} & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & p_{12} & p_{12} & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & p_{22} & p_{22} & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & p_{22} & p_{22} \end{bmatrix} \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \\ \xi_{3t} \\ \xi_{4t} \\ \xi_{5t} \\ \xi_{6t} \\ \xi_{7t} \\ \xi_{8t} \end{bmatrix}
$$

$$
= \begin{bmatrix} p_{11}\xi_{1t} + p_{11}\xi_{2t} \\ p_{11}\xi_{3t} + p_{11}\xi_{4t} \\ p_{21}\xi_{5t} + p_{21}\xi_{6t} \\ p_{21}\xi_{7t} + p_{21}\xi_{8t} \\ p_{12}\xi_{1t} + p_{12}\xi_{2t} \\ p_{12}\xi_{3t} + p_{12}\xi_{4t} \\ p_{22}\xi_{5t} + p_{22}\xi_{6t} \\ p_{22}\xi_{7t} + p_{22}\xi_{8t} \end{bmatrix}
$$

$$
= \begin{bmatrix} p_{11}\left(\xi_{1t} + \xi_{2t}\right) \\ p_{11}\left(\xi_{3t} + \xi_{4t}\right) \\ p_{21}\left(\xi_{5t} + \xi_{6t}\right) \\ p_{21}\left(\xi_{7t} + \xi_{8t}\right) \\ p_{12}\left(\xi_{1t} + \xi_{2t}\right) \\ p_{12}\left(\xi_{3t} + \xi_{4t}\right) \\ p_{22}\left(\xi_{5t} + \xi_{6t}\right) \\ p_{22}\left(\xi_{7t} + \xi_{8t}\right) \end{bmatrix}
$$

$$
= \begin{bmatrix} p_{11}\xi_{1t}^{(2)} \\ p_{11}\xi_{2t}^{(2)} \\ p_{21}\xi_{3t}^{(2)} \\ p_{21}\xi_{4t}^{(2)} \\ p_{12}\xi_{1t}^{(2)} \\ p_{12}\xi_{2t}^{(2)} \\ p_{22}\xi_{3t}^{(2)} \\ p_{22}\xi_{4t}^{(2)} \end{bmatrix} \tag{A.9}
$$

Appendix 2.B Sketch of MSM-VAR estimation alogrithm using EM optimization

MS models are typically solved by the use of the Expectation Maximization (EM) algorithm and the BLHK filter.

The EM algorithm is a heuristic procedure that allows to estimate a distribution parameters, given initial guesses about the shape of the distribution function and magnitude of its parameters. The EM algorithm is especially useful for the estimation of MS models, where, although the likelihood depends on the states, the latter are not observable, but only inferred from the data. The EM algorithm is composed of two steps: an expectation step and a maximization step.

The BLHK filter is an algorithm used in the expectation step of the EM algorithm that takes as inputs the density functions and the transition matrix and outputs a time series of expected, inferenced and smoothed state vectors.

The EM algorithm will be showcased for MSM(M) models and their derivatives. Other specifications employ the same apparatus but the state space extension.

2.B.1 Expectation step and the BLHK filter

In order to be employed, the EM algorithm requires a guess about the parameters of both the distribution and the transition probabilities. Such a guess will be the input of the expectation step of the first iteration of the alogrithm. Subsequent expectation steps will use as inputs the parameter set output of the maximization step of the precedent iteration.

Mathematically, given a primitive transition matrix P and the associated ergodic probabilities π, we set

$$\hat{\xi}_{1|0} = \otimes_{l=0}^{p} \pi \tag{B.1}$$

and

$$F = \{\text{diag}\,[\text{vec}\,(P)\,1_{M^{p-1}}]\}\,(1_M \otimes I_{M^p} \otimes 1'_M) \tag{B.2}$$

Then, the inferenced transition vectors are computed as

$$\hat{\xi}_{t|t} = \frac{\eta_t \odot \hat{\xi}_{t|t-1}}{1'_N \left(\eta_t \odot \hat{\xi}_{t|t-1}\right)}, \tag{B.3}$$

where

$$\eta_t = \begin{bmatrix} f\,(y_t|\,\theta_1,\, Y_{t-1}) \\ f\,(y_t|\,\theta_2,\, Y_{t-1}) \\ \vdots \\ f\,(y_t|\,\theta_N,\, Y_{t-1}) \end{bmatrix}, \tag{B.4}$$

$f\,(\cdot)$ being the density function of observation y_t given state n and sample Y_t, and where the predicted transition vectors can be obtained as

$$\hat{\xi}_{t+1|t} = F\hat{\xi}_{t|t}, \tag{B.5}$$

Finally, smoothed states vectors are computed as follows

$$\hat{\xi}_{t|T} = \left(F'\left(\hat{\xi}_{t+1|T} \oslash \hat{\xi}_{t+1|t}\right)\right) \odot \hat{\xi}_{t|t}. \tag{B.6}$$

Smoothed state probabilities vectors allow then for

$$\hat{\rho} = \hat{\xi}^{(2)} \oslash \left(\iota_M \otimes \hat{\xi}^{(1)}\right), \tag{B.7}$$

where $\rho = \text{vec}(P')$ and

$$\hat{\xi}^{(i)} = \sum_{t=1}^{T} \hat{\xi}_t^{(i)}, \tag{B.8}$$

keeping in mind that

$$\xi_t^{(r)} = \left(I_{M^r} \otimes 1'_{M^{p+1-r}}\right) \xi_t^{(p+1)} \tag{B.9}$$

and

$$\xi_{t-l}^{(1)} = \left(1_{M^l} \otimes I_M \otimes 1'_{M^{p-l}}\right) \xi_t^{(p+1)}. \tag{B.10}$$

2.B.2 Maximization Step - Likelihood function for EM optimization

The log-likelihood of observing the parameters λ, given the sample Y_T is

$$l\left(\lambda | Y_T,\, \lambda^{(i-1)}\right) := \int \ln\left[\mathrm{p}\left(Y,\, \xi | \lambda\right)\right] \mathrm{p}\left(Y,\, \xi | \lambda^{(i-1)}\right) \mathrm{d}\xi. \tag{B.11}$$

Then,

$$\begin{aligned}
\int \ln\left[\mathrm{p}\left(Y,\, \xi|\lambda\right)\right] \mathrm{p}\left(Y,\, \xi|\lambda^{(i-1)}\right) \mathrm{d}\xi &= \int \ln\left[\mathrm{p}\left(Y|\xi,\, \lambda\right) \Pr\left(\xi|\lambda\right)\right] \Pr\left(\xi|Y,\, \lambda^{(i-1)}\right) \\
&\quad \times \mathrm{p}\left(Y|\lambda^{(i-1)}\right) \mathrm{d}\xi \\
&= \mathrm{p}\left(Y|\lambda^{(i-1)}\right) \int \ln \mathrm{p}\left(Y|\xi,\, \lambda\right) \Pr\left(\xi|Y,\, \lambda^{(i-1)}\right) \mathrm{d}\xi \\
&\quad + \mathrm{p}\left(Y|\lambda^{(i-1)}\right) \int \ln \Pr\left(\xi|\lambda\right) \Pr\left(\xi|Y,\, \lambda^{(i-1)}\right) \mathrm{d}\xi,
\end{aligned} \tag{B.12}$$

so that

$$
\begin{aligned}
l\left(\lambda|Y_T,\,\lambda^{(i-1)}\right) \propto & \sum_{t=1}^{T}\sum_{\xi_t}\ln\left[\mathrm{p}\left(y_t|\xi_t,\,Y_{t-1},\,\theta\right)\right]\Pr\left(\xi_t|Y_T,\,\lambda^{(i-1)}\right) \\
& +\sum_{t=1}^{T}\sum_{\xi_{t-1}}\ln\left[\mathrm{p}\left(\xi_t|\xi_{t-1},\,\rho\right)\right]\Pr\left(\xi_t,\,\xi_{t-1}|Y_T,\,\lambda^{(i-1)}\right). \qquad \text{(B.13)}
\end{aligned}
$$

Focusing only on the estimation of the VAR parameter vector θ allows to consider only

$$
l\left(\theta|Y_T\right) \propto \sum_{t=1}^{T}\sum_{\xi_t}\ln\left[\mathrm{p}\left(y_t|\xi_t,\,Y_{t-1},\,\theta\right)\right]\Pr\left(\xi_t|Y_T,\,\lambda^{(i-1)}\right), \qquad \text{(B.14)}
$$

so that, by using the normality of the conditional densities

$$
\mathrm{p}\left(y_t|s_t=n,\,Y_{t-1},\,\theta\right) = (2\pi)^{-\frac{K}{2}}\left|\Sigma_n\right|^{-\frac{1}{2}}\exp\left\{-\frac{1}{2}u_{nt}\left(\gamma\right)'\Sigma_n^{-1}u_{nt}\left(\gamma\right)\right\}, \qquad \text{(B.15)}
$$

given the smoothed state probabilities vectors estimated in the expectation step, it is possible to write

$$
\begin{aligned}
l\left(\theta|Y_T\right) & \propto \text{const}-\frac{1}{2}\sum_{t=1}^{T}\sum_{n=1}^{N}\hat{\xi}_{nt|T}\left\{K\ln\left(2\pi\right)+\ln\left|\Sigma_n\right|+u_{nt}\left(\gamma\right)'\Sigma_n^{-1}u_{nt}\left(\gamma\right)\right\} \\
& \propto \text{const}-\frac{1}{2}\sum_{n=1}^{N}\left\{\hat{T}_n\ln\left|\Sigma_n\right|+\sum_{t=1}^{T}u_{nt}\left(\gamma\right)'\left(\hat{\xi}_{nt|T}\Sigma_n^{-1}\right)u_{nt}\left(\gamma\right)\right\} \\
& \propto \text{const}-\frac{1}{2}\sum_{n=1}^{N}\left\{\hat{T}_n\ln\left|\Sigma_n\right|+u_n\left(\gamma\right)'\left(\hat{\Xi}_n\otimes\Sigma_n^{-1}\right)u_n\left(\gamma\right)\right\} \\
& \propto \text{const}-\frac{1}{2}\sum_{n=1}^{N}\hat{T}_n\ln\left|\Sigma_n\right|+u\left(\gamma\right)'W^{-1}u\left(\gamma\right), \qquad \text{(B.16)}
\end{aligned}
$$

where

$$\hat{T}_n = \text{tr}\left(\hat{\Xi}_n\right) = 1_T'\hat{\xi}_n, \tag{B.17}$$

$$\hat{\Xi}_n = \text{diag}\left(\xi_{n1}, \ldots, \xi_{nT}\right), \tag{B.18}$$

$$W^{-1} = \begin{bmatrix} \hat{\Xi}_1 \otimes \Sigma_1^{-1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \hat{\Xi}_N \otimes \Sigma_N^{-1} \end{bmatrix}, \tag{B.19}$$

$$u(\gamma) = \begin{bmatrix} u_1(\gamma) \\ \vdots \\ u_N(\gamma) \end{bmatrix} = 1_N \otimes y - X\gamma \tag{B.20}$$

and

$$N = M^{(p+1)}. \tag{B.21}$$

2.B.3 Maximization Step - The EM Maximum Likelihood Estimator

At each iteration of the EM algorithm, the maximization step takes as input the smoothed states vector and the transition probabilities from the expectation step and use them to produce a new set of maximum likelihood parameters.

MSM Models Since in MSM models α and μ are not independent from each other, they are computed through a recursion that needs two regression equations, both deriving from the canonical form

$$\sum_{n=1}^{N} \hat{\xi}_{nt} \left(y_t - \mu_{n0}\right) = A_1 \sum_{n=1}^{N} \hat{\xi}_{nt} \left(y_{t-1} - \mu_{n1}\right) + \ldots + A_p \sum_{n=1}^{N} \hat{\xi}_{nt} \left(y_{t-p} - \mu_{np}\right) + u_t, \tag{B.22}$$

where

$$y_t,\, \mu_{nl},\, u_t \,\in\, \mathcal{M}\left(K \times 1\right), \tag{B.23}$$

$$N = M^{(p+1)}, \tag{B.24}$$

$$A_l \,\in\, \mathcal{M}\left(K \times K\right) \tag{B.25}$$

and where ξ_{nt} is a scalar representing derived states probabilities.

MSM - Regression equation 1 The first regression equation organizes the terms in a functional form useful for the estimation of the vectorised autoregressive coefficients α. Starting from this functional form, the estimator of the vectorized autoregressive coefficients can be derived as a function of the mean parameters of the previous internal iteration.

$$\sum_{n=1}^{N} \hat{\xi}_{nt} \left(y_t - \mu_{n0}\right) = \sum_{n=1}^{N} \hat{\xi}_{nt} \left(\bar{x}^{*}_{nt} \otimes I_k\right) \alpha + \sum_{n=1}^{N} \hat{\xi}_{nt} u_{nt}, \tag{B.26}$$

where

$$\bar{x}^{*}_{nt} = \bar{x}_t - e'_n \left(L'_1 Q',\, \ldots,\, L'_p Q'\right) \,\in\, \mathcal{M}\left(1 \times MK\right), \tag{B.27}$$

$$\bar{x}_t = \left(y'_{t-1},\, \ldots,\, y'_{t-p}\right) \,\in\, \mathcal{M}\left(1 \times MK\right), \tag{B.28}$$

$$L_l = 1'_{M^l} \otimes I_M \otimes 1'_{M^{p-l}} \in \mathcal{M}\left(M \times M^{p+1}\right), \tag{B.29}$$

e_n being the n-th column of I_N and Q a $(K \times M)$ matrix of state dependent means. By setting

$$\alpha = \left[\text{vec}\,(A_1)', \, \ldots, \, \text{vec}\,(A_p)'\right]' \in \mathcal{M}\left(K^2 p \times 1\right). \tag{B.30}$$

we can write

$$\sum_{n=1}^{N} \left(\hat{\Xi}_n \otimes I_K\right) [y - 1_T \otimes \mu_n] = \left[\sum_{n=1}^{N} \left(\hat{\Xi}_n \bar{X}_n^*\right) \otimes I_K\right] \alpha + \sum_{n=1}^{N} \left(\hat{\Xi}_n \otimes I_K\right) u_n, \tag{B.31}$$

where

$$y = (y_1', \, \ldots, \, y_T')' \in \mathcal{M}\,(TK \times 1), \tag{B.32}$$

$$\bar{X}_n^* = \left(\bar{x}_{n1}^{*\prime}, \, \ldots, \, \bar{x}_{nT}^{*\prime}\right)' = \bar{X} - 1_T \otimes e_n' \left(L_1' M', \, \ldots, \, L_p' M'\right) \in \mathcal{M}\,(T \times Kp), \tag{B.33}$$

$$\bar{X} = (\bar{x}_1', \, \ldots, \bar{x}_T')' \in \mathcal{M}\,(T \times Kp), \tag{B.34}$$

$$u_n = (u_{n1}', \, \ldots, \, u_{nT}')' \in \mathcal{M}\,(TK \times 1), \tag{B.35}$$

$$\mu_n = (e_n' L_0' M')' \in \mathcal{M}\,(1 \times K), \tag{B.36}$$

$$\hat{\Xi}_n = \text{diag}\left(\hat{\xi}_n\right) \in \mathcal{M}\left(T \times T\right) \tag{B.37}$$

and

$$\hat{\xi}_n = \left(\hat{\xi}_{n1}, \ldots, \hat{\xi}_{nT}\right)' \in \mathcal{M}\left(T \times 1\right). \tag{B.38}$$

Rearranging, we can write

$$y = \sum_{n=1}^{N} \left(\hat{\xi}_n \otimes I_K\right) \mu_n + \left[\sum_{n=1}^{N} \left(\hat{\Xi}_n \bar{X}_n^*\right) \otimes I_K\right] \alpha + u. \tag{B.39}$$

MSM - Regression equation 2 The second regression equation arranges the terms in an equation suited for the contemporaneous estimation of the vectorized conditional mean coefficients μ. from this equation the estimator of the mean parameters can be expressed as a function of the autoregressive parameters of the previous internal iteration.

$$\begin{aligned} y_t &= \sum_{n=1}^{N} \hat{\xi}_{nt} \left(\mu_{n0} + A_1 \mu_{n1} + \ldots + A_p \mu_{np}\right) + \left(\bar{x}_t \otimes I_k\right) \alpha + u_t \\ &= \left(\sum_{n=1}^{N} \hat{\xi}_{nt} X_n^{\mu}\right) \mu + \left(\bar{x}_t \otimes I_K\right) \alpha + u_t, \end{aligned} \tag{B.40}$$

where

$$\mu \in \mathcal{M}\left(MK \times 1\right). \tag{B.41}$$

$$\alpha = \left[\text{vec}\left(A_1\right)', \ldots, \text{vec}\left(A_p\right)'\right]' \in \mathcal{M}\left(K^2 p \times 1\right). \tag{B.42}$$

$$X_n^{\mu} = \left(e_n' \otimes I_K\right) X^{\mu} \in \mathcal{M}\left(K \times MK\right), \tag{B.43}$$

$$X^{\mu} = -\sum_{j=0}^{p} L_l' \otimes A_l \in \mathcal{M}\left(M^{p+1}K \times MK\right), \tag{B.44}$$

$$L_l = 1_{M^l}' \otimes I_M \otimes 1_{M^{p-l}}' \in \mathcal{M}\left(M \times M^{p+1}\right), \tag{B.45}$$

$$A_0 = -I_K \tag{B.46}$$

and

$$\bar{x}_t = \left(y_{t-1}', \ldots, y_{t-p}'\right) \in \mathcal{M}\left(1 \times MK\right). \tag{B.47}$$

Then, rearranging, the second regression equation can be formulated as

$$y = \left(\sum_{n=1}^{N} \xi_n \otimes X_n^{\mu}\right) \mu + \left(\bar{X}_t \otimes I_K\right) \alpha + u \tag{B.48}$$

MSM - The estimator Consider, for the estimations that will follow, the log likelihood function

$$\begin{aligned}
l\left(\theta | Y_T\right) &\propto \text{const} - \frac{1}{2} \sum_{t=1}^{T} \sum_{n=1}^{N} \hat{\xi}_{nt|T} \left\{K \ln\left(2\pi\right) + \ln\left|\Sigma\right| + u_{nt}\left(\gamma\right)' \Sigma^{-1} u_{nt}\left(\gamma\right)\right\} \\
&\propto \text{const} - \frac{1}{2} \sum_{n=1}^{N} \left\{\hat{T}_n \ln\left|\Sigma\right| + \sum_{t=1}^{T} u_{nt}\left(\gamma\right)' \left(\hat{\xi}_{nt|T} \Sigma^{-1}\right) u_{nt}\left(\gamma\right)\right\} \\
&\propto \text{const} - \frac{1}{2} \sum_{n=1}^{N} \left\{\hat{T}_n \ln\left|\Sigma\right| + u_n\left(\gamma\right)' \left(\hat{\Xi}_n \otimes \Sigma^{-1}\right) u_n\left(\gamma\right)\right\} \\
&\propto \text{const} - \frac{1}{2} \sum_{n=1}^{N} \hat{T}_n \ln\left|\Sigma\right| + u\left(\gamma\right)' W^{-1} u\left(\gamma\right),
\end{aligned} \tag{B.49}$$

where

$$W^{-1} = \hat{\Xi} \otimes \Sigma, \tag{B.50}$$

$$\hat{\Xi} = \text{diag}\left(\hat{\xi}\right) \tag{B.51}$$

$$\hat{\xi} = \left(\hat{\xi}_1', \ldots, \hat{\xi}_N'\right)' \tag{B.52}$$

and

$$\hat{\xi}_n = \left(\hat{\xi}_{n1}, \ldots, \hat{\xi}_{nT}\right)' \tag{B.53}$$

MSM - Estimation of α

Consider

$$\begin{aligned} y &= \sum_{n=1}^{N} \left(\hat{\xi}_n \otimes I_K\right) \tilde{\mu}_n + \left[\sum_{n=1}^{N} \left(\hat{\Xi}_n \bar{X}_n^*\right) \otimes I_K\right] \alpha + \sum_{n=1}^{N} \left(\hat{\Xi}_n \otimes I_K\right) u_n \\ &= \sum_{n=1}^{N} \left(\hat{\xi}_n \otimes I_K\right) \tilde{\mu}_n + \left[\sum_{n=1}^{N} \left(\hat{\Xi}_n \bar{X}_n^*\right) \otimes I_K\right] \alpha + u, \end{aligned} \tag{B.54}$$

where $\tilde{\mu}_m, \forall m = 1, \ldots, M$ is either a guess, when initializing the EM algorithm, the result of the previous EM cycle, at the beginning of the iteration internal to the maximization step, or the result of a previous cycle of said internal iteration.

Expressing the state dependent error terms as a function of α leads to

$$\begin{aligned} u_n(\alpha) &= (y - 1_T \otimes \tilde{\mu}_n) - \left(\bar{X}_n^* \otimes I_K\right) \alpha \\ &= \bar{y}_n^* - X_n \alpha, \end{aligned} \tag{B.55}$$

so that

$$u(\alpha) = \bar{y} - X\alpha, \tag{B.56}$$

where

$$\bar{y} = \left[(y - 1_T \otimes \mu_1)', \ldots, (y - 1_T \otimes \mu_N)'\right]' = \left(\bar{y}_1^{*\prime}, \ldots, \bar{y}_N^{*\prime}\right)' \tag{B.57}$$

$$X = \left[\left(\bar{X}_1^* \otimes I_K\right)', \ldots, \left(\bar{X}_N^* \otimes I_K\right)'\right]' = \left(X_1', \ldots, X_N'\right)'. \tag{B.58}$$

Then, the first order condition for α is

$$\begin{aligned}
\frac{\partial l(\theta|Y_T)}{\partial \alpha} &= \frac{\partial}{\partial \alpha}\left[u(\alpha)' W^{-1} u(\alpha)\right] \\
&= u(\alpha)' W^{-1} \frac{\partial}{\partial \alpha} u(\alpha) \\
&= (\bar{y} - X\alpha)' W^{-1} X \\
&= 0,
\end{aligned} \tag{B.59}$$

so that

$$\begin{aligned}
(\bar{y} - X\alpha)' W^{-1} X &= \bar{y}' W^{-1} X - \alpha' X' W^{-1} X \\
&= 0
\end{aligned} \tag{B.60}$$

and, finally,

$$\tilde{\alpha} = \left(X' W^{-1} X\right)^{-1} X' W^{-1} \bar{y}$$

$$
\begin{aligned}
&= \left[\sum_{n=1}^{N} X_n' \left(\hat{\Xi}_n \otimes \tilde{\Sigma}^{-1}\right) X_n\right]^{-1} \sum_{n=1}^{N} X_n' \left(\hat{\Xi}_n \otimes \tilde{\Sigma}^{-1}\right) \bar{y}_n^* \\
&= \left[\sum_{n=1}^{N} \left(\bar{X}_n^* \otimes I_K\right)' \left(\hat{\Xi}_n \otimes \tilde{\Sigma}^{-1}\right) \left(\bar{X}_n^* \otimes I_K\right)\right]^{-1} \\
&\times \sum_{n=1}^{N} \left(\bar{X}_n^* \otimes I_K\right)' \left(\hat{\Xi}_n \otimes \tilde{\Sigma}^{-1}\right) \left(y - 1_T \otimes \tilde{\mu}_n\right) \\
&= \left[\sum_{n=1}^{N} \left(\bar{X}_n^{*\prime} \hat{\Xi}_n \bar{X}_n^*\right) \otimes \tilde{\Sigma}^{-1}\right]^{-1} \\
&\times \sum_{n=1}^{N} \left[\left(\bar{X}_n^{*\prime} \hat{\Xi}_n\right) \otimes \tilde{\Sigma}^{-1}\right] \left(y - 1_T \otimes \tilde{\mu}_n\right) \\
&= \left[\sum_{n=1}^{N} \left(\bar{X}_n^{*\prime} \hat{\Xi}_n \bar{X}_n^*\right) \otimes I_K\right]^{-1} \\
&\times \sum_{m=1}^{M} \left[\sum_{n(m)} \left(\bar{X}_n^{*\prime} \hat{\Xi}_n\right) \otimes I_K\right] \left(y - 1_T \otimes \tilde{\mu}_m\right). \qquad \text{(B.61)}
\end{aligned}
$$

Notice that $\tilde{\Sigma}$ cancels out, otherwise it would be the result of either a guess, at the beginning of the EM algorithm, or the previous maximization step of the EM algorithm.

MSM - Estimation of μ

Consider

$$
y = \left(\sum_{n=1}^{N} \xi_n \otimes \tilde{X}_n^{\mu}\right) \mu + \left(\bar{X}_t \otimes I_K\right) \tilde{\alpha} + u, \qquad \text{(B.62)}
$$

where $\tilde{\alpha}$ is either a guess, when initializing the EM algorithm, the result of the previous EM cycle, at the beginning of the iteration internal to the maximization step, or the result of a previous cycle of said internal iteration.

Expressing the state dependent error terms as a function of μ, leads to

$$
u_n(\mu) = y - \left(\bar{X} \otimes I_K\right) \alpha - \left(1_T \otimes \tilde{X}_n^{\mu}\right) \mu, \qquad \text{(B.63)}
$$

so that

$$u\left(\mu\right)=1_{N}\otimes\bar{y}-X\mu, \tag{B.64}$$

where

$$\bar{y}=y-\left(\bar{X}\otimes I_{K}\right)\alpha \tag{B.65}$$

and

$$X=\left[\left(1_{T}\otimes\tilde{X}_{1}^{\mu}\right)',\ \ldots,\ \left(1_{T}\otimes\tilde{X}_{N}^{\mu}\right)'\right]'. \tag{B.66}$$

Then, the first order condition for μ is

$$\begin{aligned}\frac{\partial l\left(\theta|Y_{T}\right)}{\partial\mu}&=\frac{\partial}{\partial\mu}\left[u\left(\mu\right)'W^{-1}u\left(\mu\right)\right]\\&=u\left(\mu\right)'W^{-1}\frac{\partial}{\partial\mu}u\left(\mu\right)\\&=\left(1_{M}\otimes\bar{y}-X\mu\right)'W^{-1}X\\&=0,\end{aligned} \tag{B.67}$$

so that

$$\begin{aligned}\left(1_{M}\otimes\bar{y}-X\mu\right)'W^{-1}X&=\left(1_{M}\otimes\bar{y}\right)'W^{-1}X-\mu'X'W^{-1}X\\&=0\end{aligned} \tag{B.68}$$

and, finally,

$$
\begin{aligned}
\tilde{\mu} &= \left(X'W^{-1}X\right) X'W^{-1}\left(1_N \otimes \bar{y}\right) \\
&= \left[\sum_{n=1}^{N}\left(1_T \otimes \tilde{X}_n^{\mu}\right)'\left(\hat{\Xi}_n \otimes \tilde{\Sigma}^{-1}\right)\left(1_T \otimes \tilde{X}_n^{\mu}\right)\right]^{-1} \\
&\times \left[\sum_{n=1}^{N}\left(1_T \otimes \tilde{X}_n^{\mu}\right)'\left(\hat{\Xi}_n \otimes \tilde{\Sigma}^{-1}\right)\right]\left[y - \left(\bar{X} \otimes I_K\right)\tilde{\alpha}\right] \\
&= \left[\sum_{n=1}^{N}\left(1_T'\hat{\Xi}_n 1_T\right) \otimes \left(\tilde{X}_n^{\mu\prime}\tilde{\Sigma}^{-1}\tilde{X}_n^{\mu}\right)\right]^{-1} \\
&\times \left[\sum_{n=1}^{N}\left(1_T'\hat{\Xi}_n\right) \otimes \left(\tilde{X}_n^{\mu\prime}\tilde{\Sigma}^{-1}\right)\right]\left[y - \left(\bar{X} \otimes I_K\right)\tilde{\alpha}\right] \\
&= \left[\sum_{n=1}^{N}\hat{T}_n \tilde{X}_n^{\mu\prime}\tilde{\Sigma}^{-1}\tilde{X}_n^{\mu}\right]^{-1} \\
&\times \left[\sum_{n=1}^{N}\hat{\xi}_n' \otimes \left(\tilde{X}_n^{\mu\prime}\tilde{\Sigma}^{-1}\right)\right]\left[y - \left(\bar{X} \otimes I_K\right)\tilde{\alpha}\right], \qquad \text{(B.69)}
\end{aligned}
$$

where $\tilde{\Sigma}$ is the result of either a guess, at the beginning of the EM algorithm, or the previous maximization step of the EM algorithm.

MSM - Estimation of Σ

Given the estimates of $\tilde{\alpha}$ and $\tilde{\mu}$, the first order condition for Σ is

$$
\begin{aligned}
\frac{\partial l\left(\theta|Y_T\right)}{\partial \Sigma} &= \frac{\partial}{\partial \Sigma}\left[-\frac{T}{2}\ln|\Sigma| + u\left(\Sigma\right)' W^{*-1} u\left(\Sigma\right)\right] \\
&= -\frac{T}{2}\Sigma^{-1} + \frac{1}{2}\Sigma^{-1}u\left(\Sigma\right)' u\left(\Sigma\right)\Sigma^{-1} \\
&= 0, \qquad \text{(B.70)}
\end{aligned}
$$

where

$$
W^{*-1} = I_T \otimes \Sigma^{-1}, \qquad \text{(B.71)}
$$

$$u(\Sigma) = \text{diag}\left[\left(\sqrt{\hat{\xi}_1'},\, \ldots,\, \sqrt{\hat{\xi}_N'}\right) \otimes I_K\right] u(\tilde{\alpha},\, \tilde{\mu}) \tag{B.72}$$

and

$$\hat{\xi}_n = \left(\hat{\xi}_{n1},\, \ldots,\, \hat{\xi}_{nT}\right)', \tag{B.73}$$

so that

$$\tilde{\Sigma} = \frac{1}{T}\sum_{n=1}^{N} u(\tilde{\alpha},\, \tilde{\mu})' \,\hat{\Xi}_n u(\tilde{\alpha},\, \tilde{\mu}). \tag{B.74}$$

MSMAH-VAR Model In MSMAH models, the internal iteration is performed in a similar fashion of the MSM models, but the functional form for the estimation of both the autoregressive coefficients and the variance covariance matrix must be adapted to make the estimators be functions of state conditional parameters. The model can be initially expressed as

$$\begin{aligned}\sum_{m=1}^{M}\sum_{n(m)} \hat{\xi}_{nt}(y_t - \mu_{m0}) &= \sum_{m=1}^{M}\sum_{n(m)} \hat{\xi}_{nt} A_{m1}(y_{t-1} - \mu_{n1}) \\ &\quad + \ldots + \sum_{m=1}^{M}\sum_{n(m)} \hat{\xi}_{nt} A_{mp}(y_{t-p} - \mu_{np}) + \sum_{m=1}^{M}\sum_{n(m)} \hat{\xi}_{nt} u_{nt},\end{aligned} \tag{B.75}$$

where

$$n(m):\; n = (m-1)\frac{N}{M} + 1,\, \ldots,\, m\frac{N}{M}. \tag{B.76}$$

MSMAH - Regression equation 1 The first regression equation allows to express the estimator of autoregressive parameters as a function of the state conditional mean parameters estimated in the previous internal iteration.

$$\sum_{m=1}^{M}\sum_{n(m)}\hat{\xi}_{nt}\left(y_t-\mu_{m0}\right)=\sum_{m=1}^{M}\sum_{n(m)}\hat{\xi}_{nt}\left(\bar{x}_{nt}^{*}\otimes I_K\right)\alpha_m+\sum_{m=1}^{M}\sum_{n(m)}\hat{\xi}_{nt}u_{nt}, \tag{B.77}$$

where

$$\bar{x}_{nt}^{*}=\bar{x}_t-e_n'\left(L_1'Q',\,\ldots,\,L_p'Q'\right)\,\in\,\mathcal{M}\left(1\times MK\right), \tag{B.78}$$

$$\bar{x}_t=\left(y_{t-1}',\,\ldots,\,y_{t-p}'\right)\,\in\,\mathcal{M}\left(1\times MK\right), \tag{B.79}$$

$$L_l=1_{M^l}'\otimes I_M\otimes 1_{M^{p-l}}'\,\in\,\mathcal{M}\left(M\times M^{p+1}\right), \tag{B.80}$$

e_n being the n-th column of I_N and Q a $(K\times M)$ matrix of state dependent means, and

$$\alpha_m=\left[\text{vec}\left(A_{m1}\right)',\,\ldots,\,\text{vec}\left(A_{mp}\right)'\right]'\,\in\,\mathcal{M}\left(K^2p\times 1\right), \tag{B.81}$$

so that

$$\alpha=\left(\alpha_1,\,\ldots,\,\alpha_M\right)\,\in\,\mathcal{M}\left(K^2p\times M\right). \tag{B.82}$$

Rearranging

$$\begin{aligned}\sum_{m=1}^{M}\sum_{n(m)}\left(\tilde{\Xi}_n\otimes I_K\right)\left[y-1_T\otimes\mu_m\right]=&\sum_{m=1}^{M}\sum_{n(m)}\left[\left(\tilde{\Xi}_n\bar{X}_n^{*}\right)\otimes I_K\right]\alpha_m\\&+\sum_{m=1}^{M}\sum_{n(m)}\left(\tilde{\Xi}_n\otimes I_K\right)u_n,\end{aligned} \tag{B.83}$$

where

$$y = \left(y_1', \, \ldots, \, y_T'\right)' \, \in \, \mathcal{M}\left(TK \times 1\right), \tag{B.84}$$

$$\bar{X}_n^* = \left(\bar{x}_{n1}^{*\prime}, \, \ldots, \, \bar{x}_{nT}^{*\prime}\right)' = \bar{X} - 1_T \otimes e_n' \left(L_1' M', \, \ldots, \, L_p' M'\right) \, \in \, \mathcal{M}\left(T \times Kp\right), \tag{B.85}$$

$$\bar{X} = \left(\bar{x}_1', \, \ldots, \bar{x}_T'\right)' \, \in \, \mathcal{M}\left(T \times Kp\right), \tag{B.86}$$

$$u_n = \left(u_{n1}', \, \ldots, \, u_{nT}'\right)' \, \in \, \mathcal{M}\left(TK \times 1\right), \tag{B.87}$$

$$\mu_n = \left(e_n' L_0' M'\right)' \, \in \, \mathcal{M}\left(1 \times K\right), \tag{B.88}$$

$$\hat{\Xi}_n = \text{diag}\left(\hat{\xi}_n\right) \, \in \, \mathcal{M}\left(T \times T\right) \tag{B.89}$$

and

$$\hat{\xi}_n = \left(\hat{\xi}_{n1}, \, \ldots, \, \hat{\xi}_{nT}\right)' \, \in \, \mathcal{M}\left(T \times 1\right), \tag{B.90}$$

so that

$$y = \sum_{m=1}^{M} \left[\sum_{n(m)} \left(\tilde{\xi}_n \otimes I_K\right)\right] \mu_m + \sum_{m=1}^{M} \left[\sum_{n(m)} \left(\tilde{\Xi}_n \bar{X}_n^*\right) \otimes I_K\right] \alpha_m + u. \tag{B.91}$$

MSMAH - Regression equation 2 The second regression is arranged to allow for the contemporaneous estimation of the conditional mean parameters.

$$
\begin{aligned}
y_t &= \sum_{n=1}^{N} \hat{\xi}_{nt} \left(\mu_{n0} + A_1 \mu_{n1} + \ldots + A_p \mu_{np}\right) + \sum_{m=1}^{M} \sum_{n(m)} \hat{\xi}_{nt} \left(\bar{x}_t \otimes I_k\right) \alpha_m + u_t \\
&= \left(\sum_{n=1}^{N} \hat{\xi}_{nt} X_n^{\mu}\right) \mu + \sum_{m=1}^{M} \left[\sum_{n(m)} \hat{\xi}_{nt} \left(\bar{x}_t \otimes I_k\right)\right] \alpha_m + u_t,
\end{aligned} \tag{B.92}
$$

where

$$\mu \in \mathcal{M}\left(MK \times 1\right), \tag{B.93}$$

$$\alpha_m = \left[\text{vec}\left(A_{m1}\right)', \ldots, \text{vec}\left(A_{mp}\right)'\right]' \in \mathcal{M}\left(K^2 p \times 1\right), \tag{B.94}$$

$$X_n^{\mu} = \left(e_n' \otimes I_K\right) X_m^{\mu}, \forall n \in n\left(m\right) \in \mathcal{M}\left(K \times MK\right), \tag{B.95}$$

$$X_m^{\mu} = -\sum_{l=0}^{p} L_l' \otimes A_{ml}, \in \mathcal{M}\left(M^{p+1} K \times MK\right), \tag{B.96}$$

$$L_l = 1_{M^l}' \otimes I_M \otimes 1_{M^{p-l}}' \in \mathcal{M}\left(M \times M^{p+1}\right), \tag{B.97}$$

$$A_{m0} = -I_K \tag{B.98}$$

and

$$\bar{x}_t = \left(y_{t-1}', \ldots, y_{t-p}'\right) \in \mathcal{M}\left(1 \times MK\right). \tag{B.99}$$

Rearranging,

$$y = \left(\sum_{n=1}^{N} \hat{\xi}_n \otimes X_n^{\mu} \right) \mu + \sum_{m=1}^{M} \left[\sum_{n(m)} \left(\hat{\Xi}_n \bar{X} \right) \otimes I_K \right] \alpha_m + u \tag{B.100}$$

MSMAH - The estimator Consider, for the estimations that will follow, the log likelihood function

$$\begin{aligned} l\left(\theta|Y_T\right) &\propto \text{const} - \frac{1}{2} \sum_{t=1}^{T} \sum_{m=1}^{M} \sum_{n(m)} \hat{\xi}_{nt|T} \left\{ K \ln\left(2\pi\right) + \ln\left|\Sigma_m\right| + u_{nt}\left(\gamma\right)' \Sigma_m^{-1} u_{nt}\left(\gamma\right) \right\} \\ &\propto \text{const} - \frac{1}{2} \sum_{m=1}^{M} \sum_{n(m)} \left\{ \hat{T}_n \ln\left|\Sigma_m\right| + \sum_{t=1}^{T} u_{nt}\left(\gamma\right)' \left(\hat{\xi}_{nt|T} \Sigma_m^{-1} \right) u_{nt}\left(\gamma\right) \right\} \\ &\propto \text{const} - \frac{1}{2} \sum_{m=1}^{M} \sum_{n(m)} \left\{ \hat{T}_n \ln\left|\Sigma_m\right| + u_n\left(\gamma\right)' \left(\hat{\Xi}_n \otimes \Sigma_m^{-1} \right) u_n\left(\gamma\right) \right\} \\ &\propto \text{const} - \frac{1}{2} \sum_{m=1}^{M} \sum_{n(m)} \hat{T}_n \ln\left|\Sigma_m\right| + u\left(\gamma\right)' W^{-1} u\left(\gamma\right), \end{aligned} \tag{B.101}$$

where

$$W^{-1} = \begin{bmatrix} \hat{\Xi}_1 \otimes \Sigma_1^{-1} & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & \hat{\Xi}_N \otimes \Sigma_N^{-1} \end{bmatrix}. \tag{B.102}$$

MSMAH - Estimation of α_m

Consider

$$y = \sum_{m=1}^{M} \left[\sum_{n(m)} \left(\hat{\xi}_n \otimes I_K \right) \right] \tilde{\mu}_m + \sum_{m=1}^{M} \left[\sum_{n(m)} \left(\hat{\Xi}_n \bar{X}_n^{*} \right) \otimes I_K \right] \alpha_m + u, \tag{B.103}$$

where $\tilde{\mu}_m$, $\forall m = 1, \dots, M$ is either a guess, when initializing the EM algorithm, the result of the previous EM cycle, at the beginning of the iteration internal to the maximization step, or the result of a previous cycle of said internal iteration.

Expressing the state dependent error terms as a function of α_m leads to

$$
\begin{aligned}
u_n(\alpha_m) &= (y - 1_T \otimes \tilde{\mu}_m) - \left(\bar{X}_n^* \otimes I_K\right)\alpha_m \\
&= \bar{y}_m^* - X_n \alpha_m,\ n \in n(m),
\end{aligned} \tag{B.104}
$$

so that

$$
u(\alpha_m) = 1_{\frac{N}{M}} \otimes \bar{y}_m^* - X_m \alpha_m, \tag{B.105}
$$

where

$$
X_m = \left(X'_{(m-1)\frac{N}{M}+1}, \ldots, X'_{m\frac{N}{M}}\right)'. \tag{B.106}
$$

Then, the first order condition for α_m is

$$
\begin{aligned}
\frac{\partial l(\theta|Y_T)}{\partial \alpha_m} &= \frac{\partial}{\partial \alpha_m}\left[u(\alpha_m)' W_m^{-1} u(\alpha_m)\right] \\
&= u(\alpha_m)' W_m^{-1} \frac{\partial}{\partial \alpha} u(\alpha_m) \\
&= (\bar{y}_m - X_m \alpha_m)' W_m^{-1} X_m \\
&= 0,
\end{aligned} \tag{B.107}
$$

where

$$
W_m^{-1} = \begin{bmatrix} \hat{\Xi}_{(m-1)\frac{N}{M}+1} \otimes \tilde{\Sigma}_m^{-1} & \ldots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \ldots & \hat{\Xi}_{m\frac{N}{M}} \otimes \tilde{\Sigma}_m^{-1} \end{bmatrix}, \tag{B.108}
$$

so that

$$
\begin{aligned}
\left(\bar{y}_m - X_m\alpha_m\right)' W_m^{-1} X_m &= \bar{y}_m' W_m^{-1} X_m - \alpha_m' X_m' W_m^{-1} X_m \\
&= 0 \qquad \text{(B.109)}
\end{aligned}
$$

and, finally,

$$
\begin{aligned}
\tilde{\alpha}_m &= \left(X_m' W_m^{-1} X_m\right)^{-1} X_m' W_m^{-1} \left(1_{\frac{N}{M}} \otimes \bar{y}_m^*\right) \\
&= \left[\sum_{n(m)} X_n' \left(\hat{\Xi}_n \otimes \tilde{\Sigma}_m^{-1}\right) X_n\right]^{-1} \left[\sum_{n(m)} X_n' \left(\hat{\Xi}_n \otimes \tilde{\Sigma}_m^{-1}\right)\right] \bar{y}_m^* \\
&= \left[\sum_{n(m)} \left(\bar{X}_n^* \otimes I_K\right)' \left(\hat{\Xi}_n \otimes \tilde{\Sigma}_m^{-1}\right) \left(\bar{X}_n^* \otimes I_K\right)\right]^{-1} \\
&\times \left[\sum_{n(m)} \left(\bar{X}_n^* \otimes I_K\right)' \left(\hat{\Xi}_n \otimes \tilde{\Sigma}_m^{-1}\right)\right] \left(y - 1_T \otimes \tilde{\mu}_m\right) \\
&= \left[\sum_{n(m)} \left(\bar{X}_n^{*\prime} \hat{\Xi}_n \bar{X}_n^*\right) \otimes \tilde{\Sigma}_m^{-1}\right]^{-1} \\
&\times \left[\sum_{n(m)} \left(\bar{X}_n^{*\prime} \hat{\Xi}_n\right) \otimes \tilde{\Sigma}_m^{-1}\right] \left(y - 1_T \otimes \tilde{\mu}_m\right) \\
&= \left[\sum_{n(m)} \left(\bar{X}_n^{*\prime} \hat{\Xi}_n \bar{X}_n^*\right) \otimes I_K\right]^{-1} \\
&\times \left[\sum_{n(m)} \left(\bar{X}_n^{*\prime} \hat{\Xi}_n\right) \otimes I_K\right] \left(y - 1_T \otimes \tilde{\mu}_m\right). \qquad \text{(B.110)}
\end{aligned}
$$

Notice that $\tilde{\Sigma_m}$ cancels out, otherwise it would be the result of either a guess, at the beginning of the EM algorithm, or the previous maximization step of the EM algorithm.

MSMAH - Estimation of μ

Consider

$$y=\left(\sum_{m=1}^{M}\sum_{n(m)}\hat{\xi}_n\otimes\tilde{X}_n^{\mu}\right)\mu+\sum_{m=1}^{M}\left[\sum_{n(m)}\left(\hat{\Xi}_n\bar{X}\right)\otimes I_K\right]\tilde{\alpha}_m+u \tag{B.111}$$

where $\tilde{\alpha}$ is either a guess, when initializing the EM algorithm, the result of the previous EM cycle, at the beginning of the iteration internal to the maximization step, or the result of a previous cycle of said internal iteration.

Expressing the state dependent error terms as a function of μ, leads to

$$u_n(\mu)=y-\left(\bar{X}\otimes I_K\right)\tilde{\alpha}_n-\left(1_T\otimes\tilde{X}_n^{\mu}\right)\mu. \tag{B.112}$$

Since

$$\tilde{\alpha}\in\mathcal{M}\left(K^2p\times M\right), \tag{B.113}$$

then

$$\tilde{\alpha}_n=\left(e_n'L_0'\tilde{\alpha}'\right)', \tag{B.114}$$

where

$$\tilde{\alpha}_n=\tilde{\alpha}_m,\ \forall n\in n(m), \tag{B.115}$$

and

$$n(m)=(m-1)\frac{N}{M}+1,\ \ldots,\ m\frac{N}{M}. \tag{B.116}$$

Then,

$$u(\mu)=\bar{y}-X\mu, \tag{B.117}$$

where

$$\bar{y} = \left\{ \left[y - \left(\bar{X} \otimes I_K \right) \tilde{\alpha}_1 \right]', \, \ldots, \, \left[y - \left(\bar{X} \otimes I_K \right) \tilde{\alpha}_N \right]' \right\}' \tag{B.118}$$

and

$$X = \left[\left(1_T \otimes \tilde{X}_1^{\mu} \right)', \, \ldots, \, \left(1_T \otimes \tilde{X}_N^{\mu} \right)' \right]'. \tag{B.119}$$

Then, the first order condition for μ is

$$\begin{aligned} \frac{\partial l \left(\theta | Y_T \right)}{\partial \mu} &= \frac{\partial l}{\partial \mu} \left[u \left(\mu \right)' W^{-1} u \left(\mu \right) \right] \\ &= u \left(\mu \right)' W^{-1} \frac{\partial l}{\partial \mu} u \left(\mu \right) \\ &= \left(\bar{y} - X \mu \right)' W^{-1} X \\ &= 0, \end{aligned} \tag{B.120}$$

so that

$$\begin{aligned} \left(\bar{y} - X \mu \right)' W^{-1} X &= \bar{y}' W^{-1} X - \mu' X' W^{-1} X \\ &= 0 \end{aligned} \tag{B.121}$$

and, finally,

$$\begin{aligned} \tilde{\mu} &= \left(X' W^{-1} X \right) X' W^{-1} \bar{y} \\ &= \left[\sum_{m=1}^{M} \sum_{n(m)} \left(1_T \otimes \tilde{X}_n^{\mu} \right)' \left(\hat{\Xi}_n \otimes \tilde{\Sigma}_m^{-1} \right) \left(1_T \otimes \tilde{X}_n^{\mu} \right) \right]^{-1} \end{aligned}$$

$$
\begin{aligned}
&\times \left\{ \sum_{m=1}^{M} \left[\sum_{n(m)} \left(1_T \otimes \tilde{X}_n^{\mu} \right)' \left(\hat{\Xi}_n \otimes \tilde{\Sigma}_m^{-1} \right) \right] \left\{ y - \left[\sum_{n(m)} \left(\hat{\Xi}_n \bar{X} \right) \otimes I_K \right] \tilde{\alpha}_m \right\} \right\} \\
&= \left[\sum_{m=1}^{M} \sum_{n(m)} \left(1_T' \hat{\Xi}_n 1_T \right) \otimes \left(\tilde{X}_n^{\mu\prime} \tilde{\Sigma}_m^{-1} \tilde{X}_n^{\mu} \right) \right]^{-1} \\
&\times \left\{ \sum_{m=1}^{M} \left[\sum_{n(m)} \left(1_T' \hat{\Xi}_n \right) \otimes \left(\tilde{X}_n^{\mu\prime} \tilde{\Sigma}_m^{-1} \right) \right] \left\{ y - \left[\sum_{n(m)} \left(\hat{\Xi}_n \bar{X} \right) \otimes I_K \right] \tilde{\alpha}_m \right\} \right\} \\
&= \left[\sum_{m=1}^{M} \sum_{n(m)} \hat{T}_n \tilde{X}_n^{\mu\prime} \tilde{\Sigma}_m^{-1} \tilde{X}_n^{\mu} \right]^{-1} \\
&\times \left\{ \sum_{m=1}^{M} \left[\sum_{n(m)} \hat{\xi}_n' \otimes \left(\tilde{X}_n^{\mu\prime} \tilde{\Sigma}_m^{-1} \right) \right] \left\{ y - \left[\sum_{n(m)} \left(\hat{\Xi}_n \bar{X} \right) \otimes I_K \right] \tilde{\alpha}_m \right\} \right\} \qquad \text{(B.122)}
\end{aligned}
$$

where $\tilde{\Sigma}_m$, $\forall m = 1, \ldots, M$ is the result of either a guess, at the beginning of the EM algorithm, or the previous maximization step of the EM algorithm.

MSMAH - Estimation of Σ_m

Given the estimates of $\tilde{\alpha}$ and $\tilde{\mu}$, the first order condition for Σ_m is

$$
\begin{aligned}
\frac{\partial l(\theta|Y_T)}{\partial \Sigma_m} &= \frac{\partial}{\partial \Sigma_m} \left\{ -\frac{1}{2} \left[\sum_{n(m)} \hat{T}_n \ln |\Sigma_m| + u(\Sigma_m)' W_m^{*-1} u(\Sigma_m) \right] \right\} \\
&= -\frac{1}{2} \left(\sum_{n(m)} \hat{T}_n \right) \Sigma_m^{-1} + \frac{1}{2} \Sigma_m^{-1} u(\Sigma_m)' u(\Sigma_m) \Sigma_m^{-1} \\
&= 0, \qquad \text{(B.123)}
\end{aligned}
$$

where

$$
W_m^{*-1} = I_T \otimes \Sigma_m^{-1}, \qquad \text{(B.124)}
$$

$$
\hat{T}_n = \text{tr}\left(\hat{\xi}_n \right), \qquad \text{(B.125)}
$$

and

$$u\left(\Sigma_m\right) = \text{diag}\left[\left(\sqrt{\hat{\xi}'_{(m-1)\frac{N}{M}+1}},\, \ldots,\, \sqrt{\hat{\xi}'_{m\frac{N}{M}}}\right) \otimes I_K\right] u\left(\tilde{\alpha}_m,\, \tilde{\mu}_m\right). \tag{B.126}$$

Finally, the estimator for Σ_m can be expressed as

$$\tilde{\Sigma}_m = \frac{1}{\hat{T}_m} \sum_{n(m)} u\left(\tilde{\alpha}_m,\, \tilde{\mu}_m\right)' \hat{\Xi}_n u\left(\tilde{\alpha}_m,\, \tilde{\mu}_m\right) \tag{B.127}$$

Appendix 2.C Proposition 2

Consider

$$\xi_t \otimes (A(s_t)\, y_{t-1}) = \begin{bmatrix} I(s_t=1)\sum_{m=1}^{M} I(s_t=m)\, A_m y_{t-1} \\ I(s_t=2)\sum_{m=1}^{M} I(s_t=m)\, A_m y_{t-1} \\ \vdots \\ I(s_t=M)\sum_{m=1}^{M} I(s_t=m)\, A_m y_{t-1} \end{bmatrix}, \tag{C.1}$$

$$\forall b = 1,\ M \Rightarrow I(s_t=b)\sum_{m=1}^{M} I(s_t=m)\, A_m y_{t-1} = \begin{cases} A_b y_{t-1}, & \text{if } I(s_t=b)=1 \\ 0, & \text{if } I(s_t=b)=0 \end{cases} \tag{C.2}$$

Then

$$\begin{aligned} \xi_t \otimes (A(s_t)\, y_{t-1}) &= \begin{bmatrix} I(s_t=1)\sum_{m=1}^{M} I(s_t=m)\, A_m y_{t-1} \\ I(s_t=2)\sum_{m=1}^{M} I(s_t=m)\, A_m y_{t-1} \\ \vdots \\ I(s_t=M)\sum_{m=1}^{M} I(s_t=m)\, A_m y_{t-1} \end{bmatrix} \\ &= \begin{bmatrix} \xi_{1t} A_1 y_{t-1} \\ \xi_{2t} A_2 y_{t-1} \\ \vdots \\ \xi_{Mt} A_M y_{t-1} \end{bmatrix} \\ &= \begin{bmatrix} \xi_{1t} \otimes (A_1 y_{t-1}) \\ \xi_{2t} \otimes (A_2 y_{t-1}) \\ \vdots \\ \xi_{Mt} \otimes (A_M y_{t-1}) \end{bmatrix} \end{aligned}$$

$$
\begin{aligned}
&= \begin{bmatrix} \left(F_{(1)}\xi_{t-1}\right) \otimes (A_1 y_{t-1}) \\ \left(F_{(2)}\xi_{t-1}\right) \otimes (A_2 y_{t-1}) \\ \vdots \\ \left(F_{(M)}\xi_{t-1}\right) \otimes (A_M y_{t-1}) \end{bmatrix} + \begin{bmatrix} v_{1t} \otimes (A_1 y_{t-1}) \\ v_{2t} \otimes (A_2 y_{t-1}) \\ \vdots \\ v_{Mt} \otimes (A_M y_{t-1}) \end{bmatrix} \\
&= \begin{bmatrix} ([p_{11} \ \dots p_{M1}]\, \xi_{t-1}) \otimes (A_1 y_{t-1}) \\ ([p_{12} \ \dots p_{M2}]\, \xi_{t-1}) \otimes (A_2 y_{t-1}) \\ \vdots \\ ([p_{1M} \ \dots p_{MM}]\, \xi_{t-1}) \otimes (A_M y_{t-1}) \end{bmatrix} + \begin{bmatrix} v_{1t} \otimes (A_1 y_{t-1}) \\ v_{2t} \otimes (A_2 y_{t-1}) \\ \vdots \\ v_{Mt} \otimes (A_M y_{t-1}) \end{bmatrix} \\
&= \begin{bmatrix} ([p_{11} \ \dots p_{M1}] \otimes A_1)\,(\xi_{t-1} \otimes y_{t-1}) \\ ([p_{12} \ \dots p_{M2}] \otimes A_2)\,(\xi_{t-1} \otimes y_{t-1}) \\ \vdots \\ ([p_{1M} \ \dots p_{MM}] \otimes A_M)\,(\xi_{t-1} \otimes y_{t-1}) \end{bmatrix} \\
&+ \begin{bmatrix} v_{1t} \otimes (A_1 y_{t-1}) \\ v_{2t} \otimes (A_2 y_{t-1}) \\ \vdots \\ v_{Mt} \otimes (A_M y_{t-1}) \end{bmatrix} \\
&= \begin{bmatrix} p_{11}A_1 & \dots & p_{M1}A_1 \\ \vdots & \ddots & \vdots \\ p_{1M}A_M & \dots & p_{MM}A_M \end{bmatrix} (\xi_{t-1} \otimes y_{t-1}) + v_t \otimes (A\,(s_t)\, y_t) \\
&= \Pi \psi_{t-1} + v_t \otimes (A\,(s_t)\, y_t)\,. \qquad \text{(C.3)}
\end{aligned}
$$

Appendix 2.D Krolzig MSMH(3)-DVAR(1) parameters

Parameters from Krolzig [27]. The equation estimated is a MSMH(3)-DVAR(1) regression, such as

$\Delta y_t = \mu(s_t) + A(\Delta y_{t-1} - \mu(s_t)) + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma(s_t))$. $\bar{\xi}$ represents the ergodic state probabilities vector.

$$A = \begin{bmatrix} 0.2611 & 0.0486 & -0.0491 & 0.2805 & 0.1756 & 0.1942 \\ 0.1353 & -0.1208 & 0.0885 & 0.0892 & -0.0438 & 0.1322 \\ 0.3405 & -0.0349 & -0.1690 & 0.0697 & -0.0619 & 0.0002 \\ 0.1466 & -0.1929 & -0.0869 & -0.1512 & 0.1849 & 0.1359 \\ 0.5093 & -0.0940 & 0.1147 & 0.2400 & -0.0883 & 0.3965 \\ 0.2214 & 0.0519 & 0.0960 & -0.0526 & 0.1102 & -0.0565 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 0.3490 & -0.3009 & -0.1210 & 0.2879 & -0.0043 & -0.1371 \\ -0.3009 & 1.1041 & 0.7852 & 0.1396 & 0.1568 & -0.0554 \\ -0.1210 & 0.7852 & 4.2340 & 0.4872 & 0.2583 & 0.3914 \\ 0.2879 & 0.1396 & 0.4872 & 1.7697 & 0.1680 & 0.4827 \\ -0.0043 & 0.1568 & 0.2583 & 0.1680 & 0.2563 & -0.0908 \\ -0.1371 & -0.0554 & 0.3914 & 0.4827 & -0.0908 & 1.2117 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 0.5479 & -0.0616 & -0.0358 & 0.1163 & 0.1942 & 0.1029 \\ -0.0616 & 0.3746 & 0.1008 & 0.3399 & 0.0955 & 0.0972 \\ -0.0358 & 0.1008 & 1.0126 & 0.5358 & 0.1223 & 0.1096 \\ 0.1163 & 0.3399 & 0.5358 & 1.5798 & 0.3222 & 0.0055 \\ 0.1942 & 0.0955 & 0.1223 & 0.3222 & 0.7137 & 0.0775 \\ 0.1029 & 0.0972 & 0.1096 & 0.0055 & 0.0775 & 1.0847 \end{bmatrix}$$

$$\Sigma_3 = \begin{bmatrix} 1.9401 & 1.2769 & 0.4678 & 0.1393 & 0.8160 & 0.6396 \\ 1.2769 & 1.5155 & 0.4951 & 0.4007 & 0.0912 & 0.1197 \\ 0.4678 & 0.4951 & 0.7174 & 0.1230 & -0.3385 & -0.4429 \\ 0.1393 & 0.4007 & 0.1230 & 0.5569 & -0.2180 & 0.1382 \\ 0.8160 & 0.0912 & -0.3385 & -0.2180 & 1.0667 & 1.0594 \\ 0.6396 & 0.1197 & -0.4429 & 0.1382 & 1.0594 & 1.7205 \end{bmatrix}$$

$$\mu_1 = \begin{bmatrix} 1.2692 \\ 2.8917 \\ 1.4081 \\ 1.2664 \\ 1.6168 \\ 1.8695 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 0.5998 \\ 1.0280 \\ 0.5785 \\ 0.7225 \\ 0.7755 \\ 0.7742 \end{bmatrix}, \quad \mu_3 = \begin{bmatrix} -0.4463 \\ 0.7623 \\ 0.3866 \\ -0.8935 \\ -0.1095 \\ -0.0551 \end{bmatrix}$$

$$P = \begin{bmatrix} 0.9213 & 0.0786 & 0.0001 \\ 0.0287 & 0.8418 & 0.1295 \\ 0.0000 & 0.4148 & 0.5852 \end{bmatrix}$$

$$\bar{\xi} = \begin{bmatrix} 0.2178 \\ 0.5961 \\ 0.1861 \end{bmatrix}$$

Appendix 2.E GOP MSIH(3)-VAR(1) parameters

Parameters from Guidolin et al. [18]. The equation estimated is a MSIH(3)-VAR(1) regression, such as

$y_t = \nu(s_t) + Ay_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \Sigma(s_t))$. $\bar{\xi}$ represents the ergodic state probabilities vector and P is the transition matrix.

$$A = \begin{bmatrix} 0.693 & 0.066 & 0.270 & 0.201 & -0.056 & -0.003 & 0.005 & -0.008 \\ 0.228 & 0.87 & 0.092 & 0.127 & -0.039 & 0.006 & 0.043 & 0.068 \\ 0.009 & 0.004 & 0.823 & 0.062 & 0 & -0.002 & -0.007 & -0.009 \\ -0.009 & -0.001 & 0.153 & 0.601 & 0.001 & 0.001 & 0.006 & 0.006 \\ 0.053 & 0.032 & -0.244 & 0.193 & 0.785 & -0.053 & -0.035 & -0.008 \\ 0.042 & -0.028 & 0.067 & -0.357 & 0.278 & 1.058 & 0.043 & -0.008 \\ 0.004 & -0.161 & 0.701 & -0.325 & -0.033 & 0.02 & 0.981 & 0.090 \\ -0.039 & 0.191 & -0.843 & 0.605 & -0.006 & 0.007 & -0.029 & 0.872 \end{bmatrix}$$

$$\Sigma_1 = \begin{bmatrix} 0.0207 & 0.0056 & 0.0055 & 0.0077 & 0.0032 & 0.0009 & 0.0011 & 0.0007 \\ 0.0056 & 0.0069 & 0.0067 & 0.0034 & 0.0024 & 0.0001 & 0.0003 & 0.0003 \\ 0.0055 & 0.0067 & 0.2884 & 0.0299 & -0.0019 & 0.0007 & 0.0044 & 0.0036 \\ 0.0077 & 0.0034 & 0.0299 & 0.0729 & 0.0025 & -0.0014 & -0.0024 & -0.0025 \\ 0.0032 & 0.0024 & -0.0019 & 0.0025 & 0.0151 & 0.0014 & 0.0000 & -0.0006 \\ 0.0009 & 0.0001 & 0.0007 & -0.0014 & 0.0014 & 0.0035 & 0.0043 & 0.0033 \\ 0.0011 & 0.0003 & 0.0044 & -0.0024 & 0.0000 & 0.0043 & 0.0076 & 0.0063 \\ 0.0007 & 0.0003 & 0.0036 & -0.0025 & -0.0006 & 0.0033 & 0.0063 & 0.0058 \end{bmatrix}$$

$$\Sigma_2 = \begin{bmatrix} 0.8593 & 0.2285 & 2.2585 & 0.9105 & -0.0658 & -0.0204 & -0.0210 & -0.0165 \\ 0.2285 & 0.1521 & 0.9331 & 0.1350 & -0.0135 & 0.0055 & 0.0093 & 0.0094 \\ 2.2585 & 0.9331 & 19.1319 & 5.7417 & -0.1530 & -0.0047 & 0.0043 & 0.0766 \\ 0.9105 & 0.1350 & 5.7417 & 9.7219 & -0.0235 & -0.0511 & -0.1337 & -0.1194 \\ -0.0658 & -0.0135 & -0.1530 & -0.0235 & 0.1069 & 0.0375 & 0.0122 & 0.0051 \\ -0.0204 & 0.0055 & -0.0047 & -0.0511 & 0.0375 & 0.0324 & 0.0171 & 0.0095 \\ -0.0210 & 0.0093 & 0.0043 & -0.1337 & 0.0122 & 0.0171 & 0.0266 & 0.0219 \\ -0.0165 & 0.0094 & 0.0766 & -0.1194 & 0.0051 & 0.0095 & 0.0219 & 0.0240 \end{bmatrix}$$

$$\Sigma_3 = \begin{bmatrix} 0.0090 & 0.0051 & 0.0090 & -0.0116 & -0.0001 & 0.0001 & -0.0005 & -0.0010 \\ 0.0051 & 0.0106 & 0.0188 & 0.0021 & -0.0001 & -0.0000 & -0.0008 & -0.0007 \\ 0.0090 & 0.0188 & 2.5154 & 0.2534 & -0.0013 & 0.0006 & -0.0075 & -0.0082 \\ -0.0116 & 0.0021 & 0.2534 & 1.1342 & 0.0014 & -0.0017 & -0.0084 & -0.0075 \\ -0.0001 & -0.0001 & -0.0013 & 0.0014 & 0.0005 & 0.0001 & -0.0003 & -0.0002 \\ 0.0001 & -0.0000 & 0.0006 & -0.0017 & 0.0001 & 0.0004 & 0.0013 & 0.0011 \\ -0.0005 & -0.0008 & -0.0075 & -0.0084 & -0.0003 & 0.0013 & 0.0090 & 0.0087 \\ -0.0010 & -0.0007 & -0.0082 & -0.0075 & -0.0002 & 0.0011 & 0.0087 & 0.0102 \end{bmatrix}$$

$$\nu_1 = \begin{bmatrix} 0.193 \\ 0.227 \\ 1.166 \\ 1.892 \\ -0.148 \\ -0.073 \\ -0.198 \\ -0.140 \end{bmatrix}, \quad \nu_2 = \begin{bmatrix} -0.365 \\ 0.152 \\ 0.376 \\ 0.975 \\ -0.070 \\ -0.014 \\ -0.101 \\ -0.069 \end{bmatrix}, \quad \nu_3 = \begin{bmatrix} 0.035 \\ 0.196 \\ -0.003 \\ 0.017 \\ -0.103 \\ 0.015 \\ -0.077 \\ -0.057 \end{bmatrix}$$

$$\mu_1 = \begin{bmatrix} 1.349 \\ 3.585 \\ 6.866 \\ 7.106 \\ 0.183 \\ 0.220 \\ 0.562 \\ 1.546 \end{bmatrix}, \quad \mu_2 = \begin{bmatrix} 4.453 \\ 6.861 \\ 29.016 \\ 17.022 \\ -5.343 \\ -3.793 \\ -0.352 \\ 1.489 \end{bmatrix}, \quad \mu_3 = \begin{bmatrix} 5.456 \\ 6.024 \\ 8.992 \\ 7.773 \\ 4.388 \\ 4.644 \\ 4.840 \\ 5.047 \end{bmatrix}$$

$$P = \begin{bmatrix} 0.992 & 0.008 & 0.000 \\ 0.059 & 0.832 & 0.109 \\ 0.000 & 0.029 & 0.971 \end{bmatrix}$$

$$\hat{\xi} = \begin{bmatrix} 0.6078 \\ 0.0824 \\ 0.3098 \end{bmatrix}$$

Figure 2.1: **Simulated path of the Markov chain (Krolzig)**
Panels show activation of state one, two and three respectively as simulated from a Markov chain using Krolzig [27] parameters.

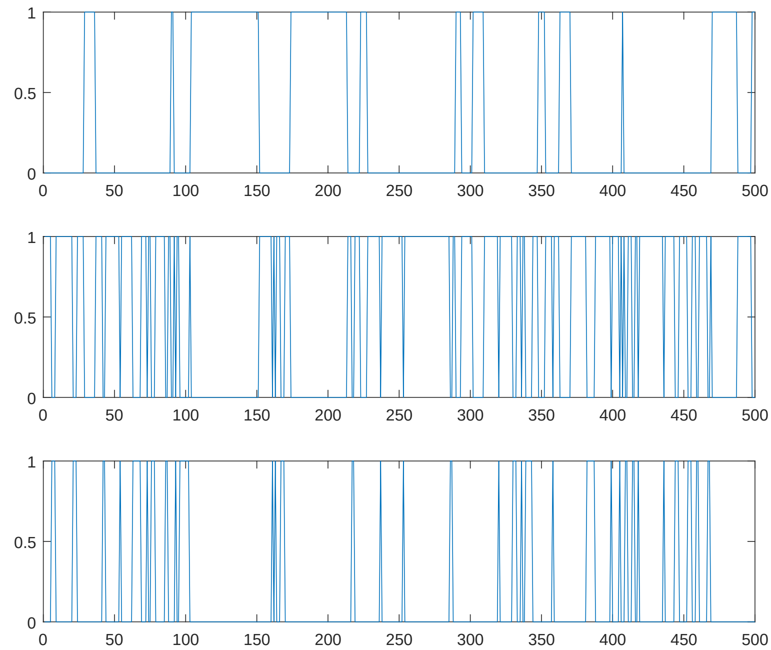


Figure 2.2: **Simulated path of the time series Krolzig**

Simulated paths under Krolzig [27] parametrization for real GNP of USA, Japan and West Germany and the real GDP of United Kingdom, Canada and Australia.

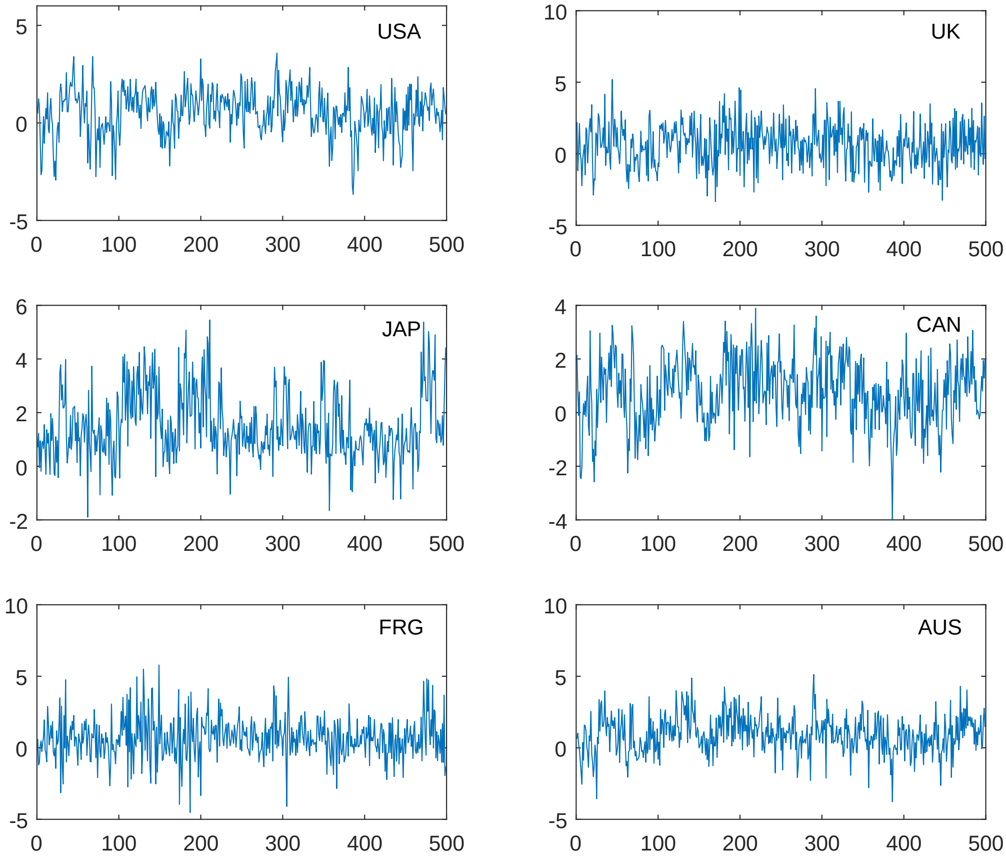


Figure 2.3: **Markov switching total connectedness (Krolzig)**
Total connectedness for a MSMH(3)-DVAR(1) as implied from a simulation with Krolzig [27] parameters and computed using smoothed state probabilities.

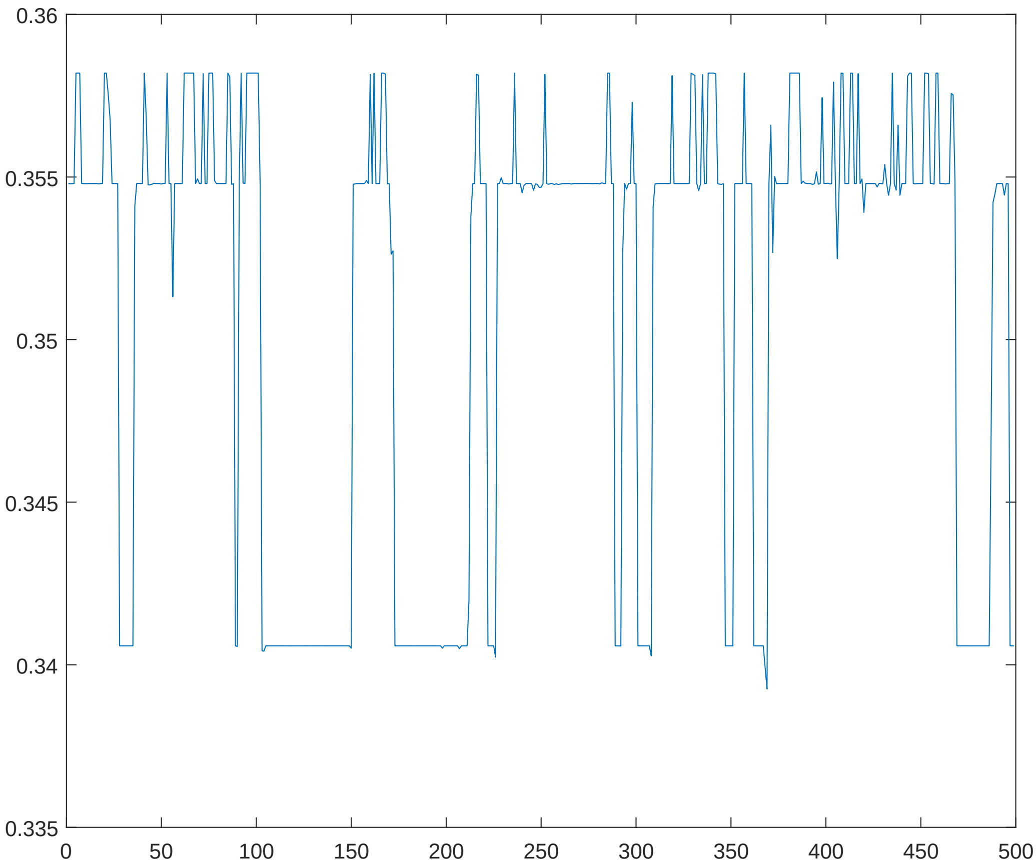


Figure 2.4: **Comparison between total connectedness and inferred state probabilities (Krolzig)**
Upper panel shows total connectedness as retrieved from the simulation with Krolzig [27] parameters. Panels two, three and four report smoothed state probabilities for state one, two and three, respectively.

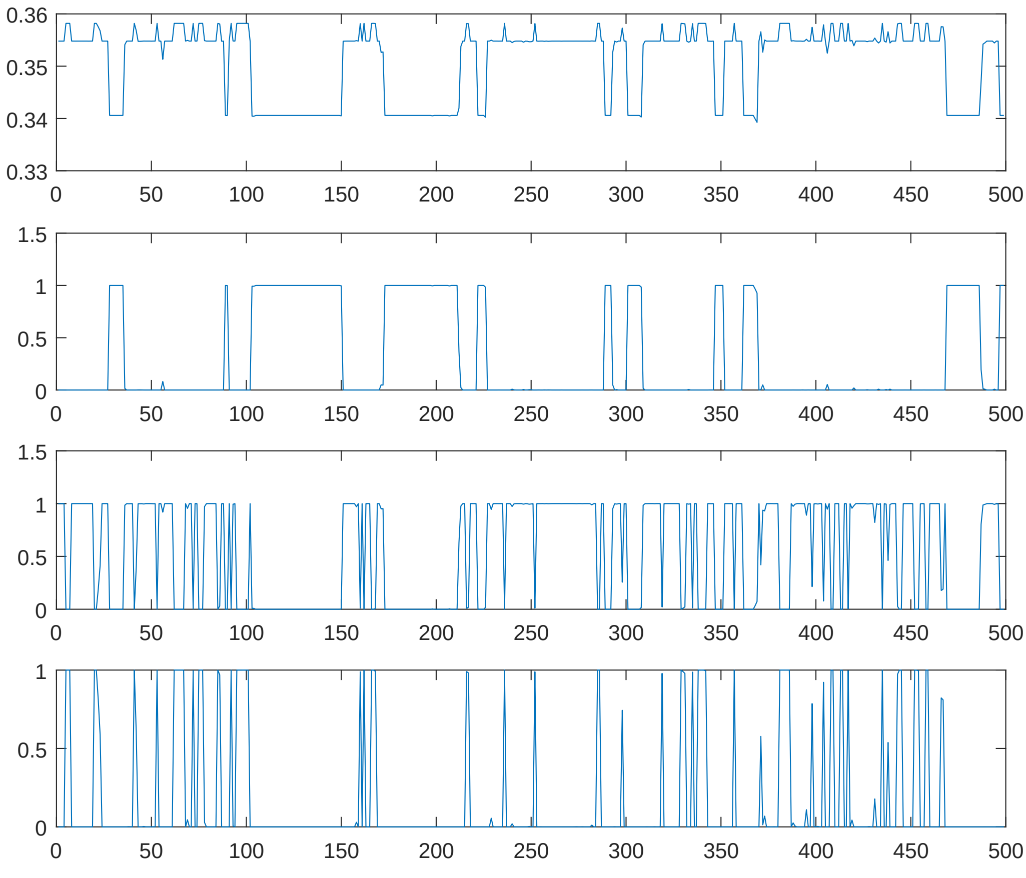


Figure 2.5: **Simulated path of the Markov chain (GOP)**
Panels show show activation of state one, two and three from a Markov chain simulation using Guidolin et al. [18] parameters.

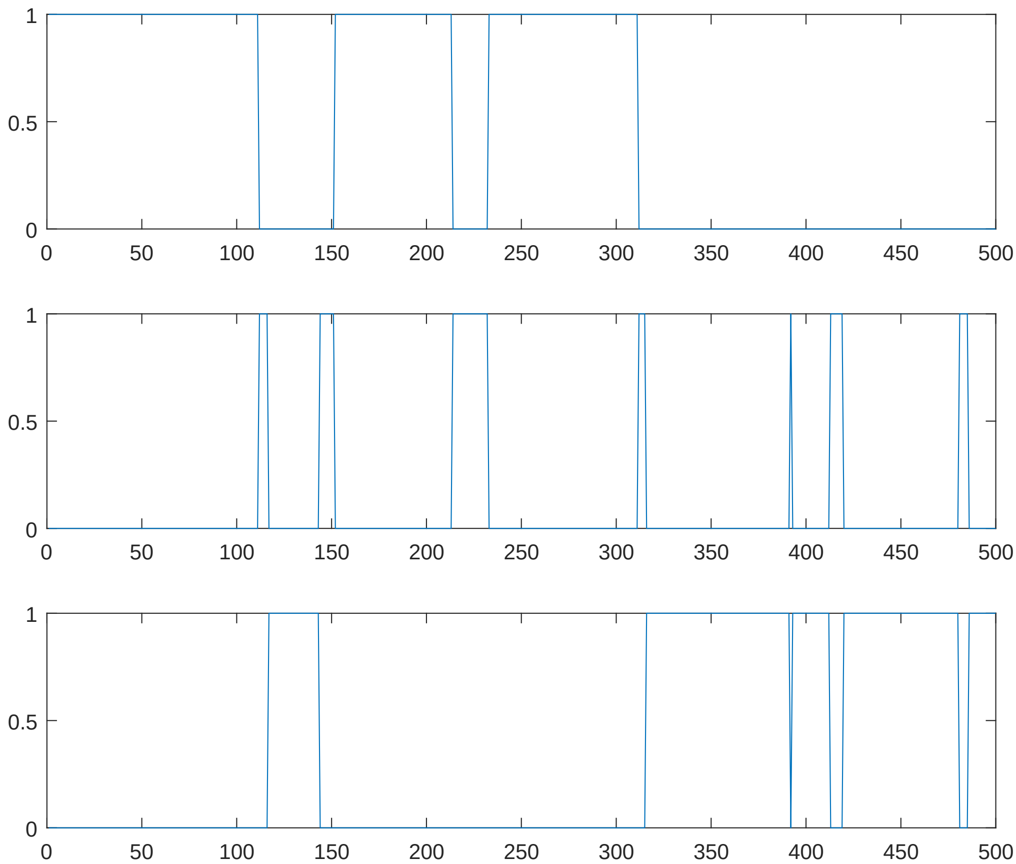


Figure 2.6: **Simulated path of the time series (GOP)**
Simulated paths, subject to Guidolin et al. [18] parameters, for investment grade short term (IGST), investment grade long term (IGLT), non investment grade short term (NIGST) and non investment grade long term (NIGLT) corporate bonds yields and one month, one year, five years and ten years Treasury bills yields.

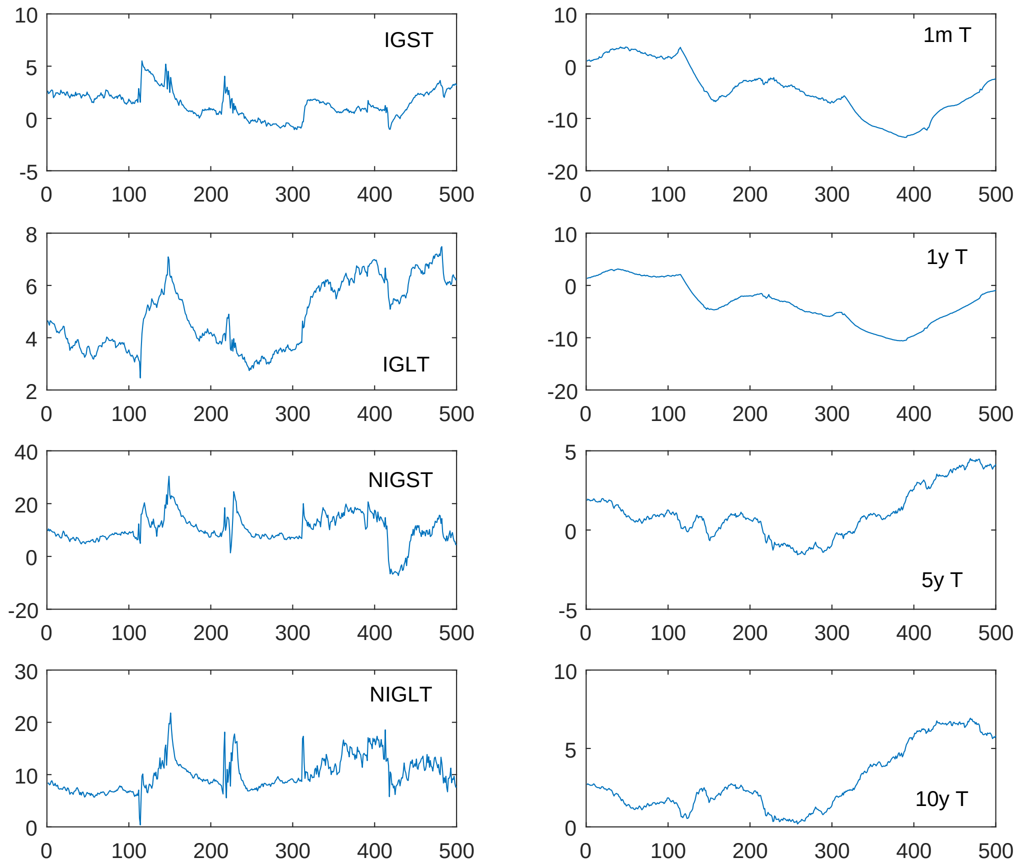


Figure 2.7: **Markov switching total connectedness (GOP)**
Total connectedness for a MSIH(3)-VAR(1) as implied from a simulation with Guidolin et al. [18] parameters and computed using smoothed state probabilities.

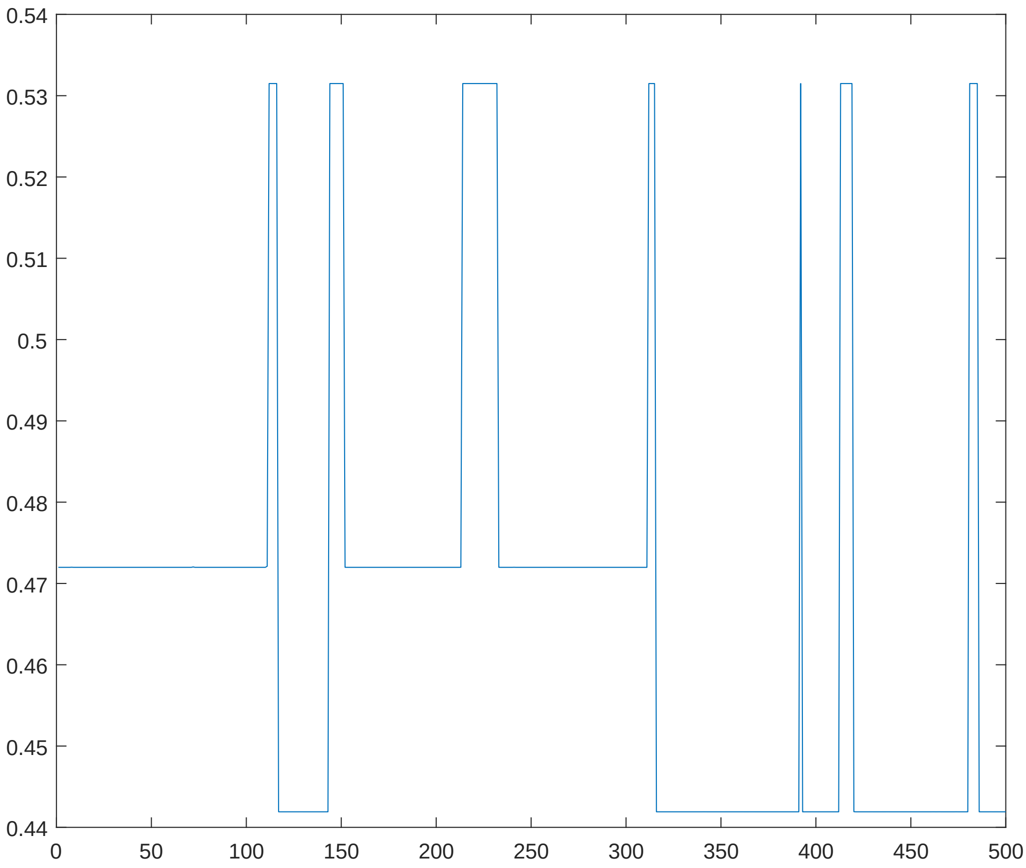


Figure 2.8: **Comparison between total connectedness and infered state probabilities (GOP)**
Upper panel shows total connectedness as retrieved from the simulation with Guidolin et al. [18] parameters. Panels two, three and four report smoothed state probabilities for state one, two and three, respectively.