Multidimensional Time Series Methods for Economics and Finance

Alessandro Celani
Matriculation no. 1092181

A thesis presented for the degree of
Doctor of Philosophy
in
Economics

Department of Economic and Social Sciences
Università Politecnica delle Marche

Supervisors:
Prof.ssa Maria Cristina Recchioni
Prof. Paolo Giudici

Science cannot solve the ultimate mystery of Nature. And that is because, in the last analysis, we ourselves are part of the mystery we are trying to solve.

Max Plank (1858 - 1947)

No problem can be solved from the same level of consciousness that created it.

Albert Einstein (1879 - 1955)

Let us try to teach generosity and altruism, because we are born selfish. Let us understand what our own selfish genes are up to, because we may then at least have the chance to upset their designs, something that no other species has ever aspired to do.

Richard Dawkins (1941 -)

Acknowledgments

The PhD drove me into an unexpected and fascinating mindset I think I'd never had without that experience.

The research activity, the new experiences, the endless conversations held with who have been with me... Rarely these conversations, started from Economics, didn't turn into something much deeper; this contributed to generating stimulant environments of deep and meaningful discussions.

That said, I deeply thank you Economics. I met you absolutely by chance, about 10 years ago, when unconsciously, and not too convinced, I enrolled at the Università Politecnica delle Marche. I didn't know what would have been unleashed.

I never thought that through Economics I would have started to inquire about the Human nature and the deep forces that move ourselves and the other species. There is no Economics without Humans; it is only the greatest and most fascinating game our species plays. It is only one of the infinite gateways through which you can enter to question about the essence of everything. Probably, the word itself does not give thanks to the depth this subject incorporates. It seems only like an ode to efficiency.

I believe that little innovation can be achieved in Economics without a profound innovation its underlying source: the Human. Without a powerful rethinking of what we believe Humans are and what should they be, it will be burdensome to end up with a genuine revolution of the Economic "Science." There is little to discuss about the causality link in this case. The game changes if we change, not vice versa. Otherwise, there will be no benevolence, generosity, altruism. All of us know these concepts; they are supposed to be the foundation of the new Social, Environmental, and Economic revolution that the world is demanding. But they don't fit so much with us right now.

Whether it is called slavery, imperialism, capitalism, communism, it doesn't

matter. Whether it is pure survival, social survival, or any other, it doesn't matter. Our nature has always been and will always be in total opposition to what we pretend to pursue. At least in the realm of the increasingly hypocritical western World.

Certain dynamics are inherently part of us, indissolubly linked to the nature of living beings. Induced by the curse of having a finite available time. Our end is marked from the exact moment we are born. And in the short journey that is life, we want to survive, at all costs, and in the best, possible way. Those dynamics are timeless and spaceless. And it applies to everyone, to thoughtful and thoughtless people. It is the foundation of the absolute incompatibility between the philosophical, utopian realm of ideas, and the raw, dystopian realm of reality.

We won't change anything if we continue to be the same specie we have always been. We won't change anything if we continue to push aside the existential questions.

By making a parallel with Economics, I would say that approaching it (only) with the Scientific Method (that gives me the job, by the way ahahah), will not solve the most puzzling issues. The dramatic objectification of our actions and thoughts, which boils down our nature to equations and equilibrium, will not give us the answer. The more we rely on elegant technicisms, the more we refine our quantitative instruments, the more we get away from the correct direction.

Hoping not to witness the n-th scientific revolution, but rather, to witness the first Human revolution!

I'm not good at giving thanks. All the people who have been, are, and will be close to me know this. There is no need to thank them here. The moments spent together, the pleasure of living my life with you, this is the best gratitude I can give to you.

And from now onward you will fill find a compendium of scientific ap-

proaches to tackle Economic problems!

Abstract

This thesis aims to address the inferential and interpretational issues in high and multi-dimensional models in the context of Economics and Finance. The growing economic and financial integration has made imperative the need to conceive Countries and Financial Markets as a single, large, interconnected entity. The main challenges induced by this framework concern the estimation and interpretation of large panels, where units can be represented by countries or assets, observed via several indicators across time.

This thesis proposes Bayesian estimation techniques for novel matrix and tensor-valued models and employs new methodological tools from Graph Theory to facilitate interpretation of high-dimensional networks. The contributions are presented in three chapters.

In Chapter 2, Graph Theory approaches are proposed to study the structures and interactions of weighted directed networks of multivariate time series observations/relationships. In Chapter 3, a Bayesian variable selection approach is proposed to handle the over-parametrization problem in large Matrix Autoregressive models. In Chapter 4, the dynamic relationship among returns, volatility, and sentiment in the cryptocurrency class is explored through a Bayesian Matrix Autoregressive model, which is the first attempt to consider financial asset data as multi-dimensional structures.

Contents

1 General introduction **1**

2 Graph Theory Measures for Panel Variance Decomposition Networks **5**

2.1 Introduction 5

2.2 Methodology 9

2.3 Data 24

2.4 Empirical results 25

2.5 Conclusion 34

3 Bayesian Stochastic Search for Matrix Autoregressive Models **36**

3.1 Introduction 36

3.2 Model and Prior Structure 39

3.3 Bayesian Estimation 45

3.4 Simulations 47

3.5 Application: Panel of Country Economic Indicators 57

3.6 Conclusion 64

4 A Multidimensional approach to Finance: on the Relationship between Sentiment, Returns and Volatility in the Crypto Market **66**

4.1 Introduction 66

4.2 Financial theory, literature and motivation 69

4.3 The Matrix Autoregressive Model 73
4.4 Real application 80
4.5 Forecasting Experiment 92
4.6 Conclusion 97

5 Concluding Remarks **99**

A Appendix **119**

A.1 Technical Details of Chapter 2 119
A.2 Technical details of Chapter 3 133
A.3 Technical details of Chapter 4 141

List of Tables

1 Country-level network hubs 33

2 Country-level network authorities 34

3 Computational time 54

4 Average MSFE and computational time 59

5 Weak exogeneity test 124

6 Diebold Mariano test P-value (MAR-AR) 147

7 Diebold Mariano test P-value (MAR-VAR) 148

8 Raftery and Lewis diagnostics 150

List of Figures

1 Global forecast error variance MST 26

2 Global forecast error variance clustering 27

3 Dependence-influence variable relationships 29

4 Country Aggregate Heatwave and Spillover index 30

5 Directed GFEVD country spillover network 32

6 Efficiency comparison between ML and MAP 51

7 Simulation performances 53

8 Tolerance 56

9 Posterior median of autoregressive coefficients 57

10 Posterior of the temporal coefficient vector 58

11 Kronecker decomposition of the GFEVD 63
12 Matrix time series of the indicators 67
13 Posterior distribution of coefficient matrices 82
14 Posterior distribution of the dynamic indicators GFEVD. . . . 83
15 Posterior of the spillover indices 84
16 Posterior of the indicator influence indices 85
17 Posterior of the indicator dependence indices 86
18 Scatter of the influence-dependence indices 87
19 Posterior distribution of the dynamic cryptos GFEVD 89
20 Posterior of the crypto influence indices 90
21 Posterior of the crypto dependence indices 91
22 Scatter of the influence-dependence cryptos 92
23 Cumulative logarithm of the RMSFE for the return indicator. 95
24 Cumulative logarithm of the RMSFE for the volatility indicator. 96
25 Cumulative logarithm of the RMSFE for the sentiment indicator. 97
26 Time series of the indicators 123
27 Fraction of significant cointegrating relationships 128
28 Directed dynamic GFEVD country spillover network 129
29 Sub-sample global forecast error variance MSTs 130
30 Sub-sample global forecast error variance clustering 131
31 Dependence-influence country relationships 132
32 Average MSE and 1 step ahead forecasting performance 139

33 One step ahead MSFE and computational time 140

34 Generalized impulse response functions for the four indicators. 142

35 Rolling BIC 143

36 Rolling cumulative logarithm of the RMSFE for the return indicator and the five different models. 144

37 Rolling cumulative logarithm of the RMSFE for the volatility indicator and the five different models. 145

38 Rolling cumulative logarithm of the RMSFE for the sentiment indicator and the five different models. 146

1 General introduction

Globalization and financial integration have substantially increased the interconnected nature of economies entities the global economy. Additionally, global catastrophes like the Great Recession and the more recent Covid-19 epidemic have highlighted how crucial it is to look at the world's systems as a web of interdependent entities. As part of the same, large, increasingly interconnected dynamic system, economies are always more susceptible to shock arising elsewhere in the world. As a consequence, statistical and econometric frameworks to estimate and thereby interpret complex high and multi-dimensional systems are always more required.

As modern data collection potential has led to accumulation of vast amount of data over time, high and multi dimensional time series are becoming more available. Among applied sciences, Economics is undoubtedly not exempted from this process. As a matter of fact, intra-country, regional-level panels (Fosten and Greenaway-McGrevy, 2022), global macroeconomic panels (Pesaran et al., 2004), as well as multilayer networks (Aldasoro and Alves, 2018) are regularly generated over time. The common thread linking these data structures is that they lie at the intersection of two or more dimensions. In the case of the first two ones, we have different indicators collected for a number of units: regions for the former, countries for the latter. Network data instead, generalize the structure just mentioned, encompassing at least a third dimension. In this framework, bilateral economic and financial flows are observed for multiple units, usually institutions or countries. Albeit to a lesser extent, also the Financial realm has been affected by this revolution. For instance, it is now common to observe various financial characteristics of many companies for different temporal horizons (Wang et al., 2019). By relaxing the assumption that units can be just institutions, we can conceive them as assets, each described by multiple characteristics. For instance, these can be represented by its price return, a proxy of the volatility, and a sentiment indicator (Celani and Pagnottoni, 2023).

Despite the above mentioned information sets may offer non negligible advantages for the understanding of the Economy, they don't come without a pitfall. In fact, their unobserved interaction can be uncovered only by employing multivariate statistics and econometric methods. In the time series framework, the workhorse is the Vector Autoregression (VAR) model (Sims, 1980).

As the number of variables to be modeled grow, the number of parameters to estimate increases exponentially. Moreover, given the intrinsic low frequency of macroeconomic data, the number of observations is usually relatively low. These two factors combined render standard frequentist estimation techniques unfeasible. In Finance the situation is just slightly better. The availability of high-frequency time series might have the potential to cope with the overparametrization issue. Nevertheless,the non-linearities induced by the highly stochastic framework usually force researchers to focus on lower dimensional sub-samples.

It is well known that modeling high dimensional time series is a delicate task, which has called for the emergence of specific literatures. In a nutshell, hierarchical regularization and factor methods are the leading approaches. Nonetheless, even when the curse of dimensionality is mitigated and a given model is (presumably) consistently estimated, the output interpretation is cumbersome. As a remedy, directed network measures based on the decomposition of the forecast error variance of a VAR have been proposed (Diebold and Yilmaz, 2014). Despite being suitable to draw a superficial picture of the network charateristics, those measures are fruitless to investigate deeper substructures. In syntheses, both inferential and interpretational threats emerges in this context.

For both of these, the additional challenges imposed by the multi dimensional setting magnify even more the already thorny framework. As such, new modeling paradigms are required to cope with these two problems simultaneously.

This thesis aims to contribute to the existing literature by addressing inferential and interpretational issues in high and multi dimensional models by:

- proposing Bayesian estimation techniques for novel matrix and tensor valued models, designed for multi dimensional data. The Bayesian approach has proven to work well as a dimensionality reduction technique. By casting it into the matrix valued framework, it has the potential to contemporaneously manage issues arising from the high and multidimensional setting, while preserving interpretative power.
- employing new methodological tools stemming from Graph Theory to facilitate interpretation of high-dimensional networks. Networks analysis represents an useful tool to describe relationships among entities and it has been extensively used in recent years to study financial and economic connectedness.

The contributions are presented in three self-contained chapters.

In Chapter 2 we propose novel Graph Theory approaches to study the structures and interactions of weighted directed networks of multivariate time series observations/relationships. Those methods are well suited to analyze connectedness matrices derived from variance decompositions of high dimensional multi-country econometric models. We apply the methodology to a number of monthly indicators of a set of European countries and determine which are the leader and the follower in the shock transmission mechanisms.

In Chapter 3 we propose a Bayesian variable selection approach to handle the over-parametrization problem in large Matrix Autoregressive (MAR) models. Two computational methods are derived: a classical Markov Chain Monte Carlo (MCMC) and a scalable Expectation Maximization (EM), the latter conceived for high dimensional settings where the computational cost of the former makes it unfeasible to estimate. Theoretical properties, com-

parative performances and computational efficiency of our proposed models are investigated through simulated and real examples.

In Chapter 4 we explore the dynamic relationship among returns, volatilty, and sentiment of the cryptocurrency class by means of a Bayesian MAR. This is the first attempt to conceive financial asset data as potential multidimensional structures. After having outlined the financial theoretical framework, we derive the frequentist and Bayesian estimator for the MAR. Eventually, we apply the proposed method to study the multidimensional relationship among returns, volalatility and sentiment of four cryptocurrencies.

Chapter 2 has been accepted for publication as: Celani, A., Cerchiello P., and Pagnottoni, P. (2022), "The Topological Structure of Panel Variance Decomposition Networks". *Journal of Financial Stability*, forthcoming. *Available at SSRN*.

Chapter 3 is currently under review process as: Celani, A., Jones, G., and Pagnottoni, P. (2022), "Bayesian Variable Selection for Matrix Autoregressive Models". *Available at SSRN*.

Chapter 4 is currently under review process as: Celani, A., and Pagnottoni, P. (2023), "The Multidimensional Relationship between Sentiment, Returns and Volatility". *Available at SSRN*.

2 Graph Theory Measures for Panel Variance Decomposition Networks

Based on the paper: Celani, A., Cerchiello P., and Pagnottoni, P. (2022), "The Topological Structure of Panel Variance Decomposition Networks". *Journal of Financial Stability*, forthcoming. *Available at SSRN*.

2.1 Introduction

Globalisation has considerably increased the interconnected nature of entities within the global economy and, thereby, their sensitivity to large shocks over the last few decades. Some of the benefits of globalisation in terms of liberalisation and development were contrasted by the emergence of the Global Financial Crisis, as a source of systemic risks posed by the financial globalisation (Mishkin, 2011). Macroeconomic linkages feature different aspects of connectedness: trade linkages, financial linkages, and price changes (Dees et al., 2007). Hence, macroeconomic connectedness is a multidimensional concept, whose fragile equilibrium can be compromised by several risk sources. Extant literature has investigated the co-movements and support the notion of global business cycles, as drivers of co-movement in national business cycles - see Kose et al. (2003, 2008). While research on systemic risk linkages and financial connectedness has considerably advanced since then, a deep understanding of international macroeconomic linkages is still a basic open question, and the network structure of shock propagation across global economies is a ground to explore.

Large, unexpected exogenous shocks, such as the COVID-19 pandemic exert dramatic impacts on the financial and economic structure of countries (Delis et al., 2021; Pagnottoni et al., 2021; Iwanicz-Drozdowska et al., 2021; Spelta and Pagnottoni, 2021; Liu et al., 2021; Ahelegbey et al., 2022; Bitetto et al., 2021; Bartolucci et al., 2021). Recent groundwork in the financial con-

text shows that the reaction of financial markets to exogenous shocks, such as climate related ones among the others, is increasingly harsher over recent years. Dafermos et al. (2018) shows how climate change has a significant effect on financial stability, as it causes a sudden drop in the level of liquidity injected to firms, as well as it reduces corporate bond prices, together with credit supply. Battiston et al. (2017) proposed a network-based climate value at risk to study risk transmission both system-wide and at institutional level, giving birth to a flourishing literature investigating the nexus between climate change and finance (Stolbova et al., 2018; Monasterolo et al., 2017; Battiston et al., 2019; Roncoroni et al., 2021; Battiston et al., 2021; Pagnottoni et al., 2022).

The modelling paradigm of current research in business cycle stems from methodologies which are not grounded on network theory - see Diebold and Yilmaz (2015) - although they present many complementary aspects to the systemic risk and to the complex network literature. Financial network econometric models include those developed by Billio et al. (2012) and Diebold and Yilmaz (2009, 2012, 2014), in which individual entities can be represented as nodes in a financial network. While Billio et al. (2012) develop causal networks based on the concept of Granger causality, Diebold and Yilmaz (2014) build generalized forecast error variance decomposition (GFEVD) on approximating Vector Autoregressive (VAR) models. Since then, the literature has started integrating time series econometrics approaches with network tools to assess systemic risk, particularly in financial market contexts - see e.g. Ahelegbey et al. (2016); Ahelegbey and Giudici (2022). More recently, Greenwood-Nimmo et al. (2021) developed a method to employ forecast error variance decomposition to evaluate the macroeconomic connectedness in any multi-country macroeconomic model with an approximate multi-country VAR representation. Therefore, such framework: a) it is convenient in terms of interpretability; b) it is able to take into account for contemporaneous impacts; c) it can predictively measure system-wide and pairwise connectedness of time series networks.

Despite having the merit of being robust from a methodological perspective, state-of-the-art econometric models - based on forecast error variance decomposition - lack of a comprehensive framework in terms of their interpretation from a network analysis viewpoint. Diebold and Yilmaz (2014) and Greenwood-Nimmo et al. (2021) are key studies in this stream of research, as they relate network systemic risk measures based on forecast error variance decomposition, although without a strong link to network theory or topological analysis. The use of robust statistical methods to measure macroeconomic interconnectedness and systemic risk, combined with complex system methods - see, e.g. Battiston et al. (2016); Bardoscia et al. (2017); Roukny et al. (2018) - is therefore a promising avenue for future research.

Against this background, we propose a network-based framework to investigate systemic risk, macroeconomic connectedness and lead-lag relationships from a set of multi-country and global macroeconometric models, i.e. Global Vector Autoregressive (GVAR) models. Our approach takes root from the forecast error variance decomposition spillover measures (Diebold and Yilmaz, 2014; Greenwood-Nimmo et al., 2021). We analyze the estimates of the econometric model by means of graph theory, studying clustering and hub-authority dynamics of the GVAR spillover network and its implications in terms of risk propagation mechanisms. This is done by means of the Minimum Spanning Tree (MST), the Louvain community detection and the Kleinber algorithms, so to derive the backbone structure, communities and hub-authority dynamics of macroeconomic relationships. To this end, indices based on the decomposition of the variance-covariance matrix of the forecast error derived from the dynamic GVAR estimates are used, so to statistically and multidimensionally capture systemic risk in a predictive framework. We apply our methodology to study the spillover network of real economy and sentiment indicators of 12 European countries over the period 01/2000-11/2021 at the global, country and variable levels.

Similarly to our approach, Elhorst et al. (2021) bring together the spatial and GVAR classes of econometric models. Motivated by the study of

spillovers, they define a measurable concept of spillover in this context based on impulse responses. Gross and Kok (2013) define a mixed-cross-section GVAR for forecasting and conducting systematic shock simulation. This eases the definition of spillover potential measures for within and across groups of sovereign and banks, also employing network centrality measures. In contrast, we examine the topological structure of GVAR spillover networks and its implications from a risk management perspective. In particular we study: a) the shortest paths of contagion; b) the clusters of shock transmission; and c) the role of nodes in the risk transmission channels. Additionally, we estimate a Bayesian GVAR, which encompasses also the maximum likelihood estimator as a special case, in order to cope with the possible presence of dominant units.

The contribution of this chapter is twofold. From a methodology perspective, we extend the econometric GFEVD method illustrated by Greenwood-Nimmo et al. (2021) to a spillover network topology framework, which enables to study the complex dynamic network structure of the time series of forecast error variance shares, i.e. of spillovers. The combination of a robust global econometric model and the concepts of network theory results in a practical method for the development of a sound set of statistical network systemic risk indicators, which are based upon predictive directional measurement of spillovers. From an empirical viewpoint, our strategy allows to simultaneously examine: a) the dynamic relationships of real economy and sentiment at a country level, identifying lead-lag relationships across EU member states over time; b) the dynamic connectedness existing between real economy and sentiment variables.

The remainder of this chapter is organized as follows. Section 2.2 introduces the proposed methodology. Section 2.3 describes the data used in our empirical study and conducts preliminary data analysis. Section 2.4 illustrates our empirical outcomes. Section 2.5 offers a final discussion.

2.2 Methodology

The Global VAR model

Global VAR models are built on the premise that simple VAR models do well describe the linear dynamics of a system, although they suffer from the overparameterization problem, which makes their use impractical in high-dimensional contexts. Among the different approaches proposed in the literature to solve the curse of dimensionality, the GVAR proposed by Pesaran et al. (2004) and surveyed in Chudik and Pesaran (2016) provides a good framework to model a relative small set of variables shared by a fixed number of spatial units - which, in our case, represent countries.

The GVAR solves the dimensionality problem by decomposing the underlying large dimensional VAR into a smaller number of conditional models, which are linked together via cross sectional averages.

The first step consists of estimating small-scale country models enlarged by weakly exogenous and possibly global variables (VARX* model). Let $\mathbf{x}_{it}$ be a k_i dimensional vector of endogenous variables of country $i = 1, ..., N$. We model each $\mathbf{x}_{it}$ as a VARX*(p_1, p_2):

$$\mathbf{x}_{it} = \boldsymbol{\alpha}_{i0} + \sum_{s=1}^{p_1} \boldsymbol{\Phi}_{is}\mathbf{x}_{i,t-s} + \sum_{r=0}^{p_2} \boldsymbol{\Lambda}_{ir}\mathbf{x}^*_{it-r} + \boldsymbol{\varepsilon}_{it}, \tag{2.1}$$

where $\boldsymbol{\alpha}_{i0}$ is a vector of intercepts, $\boldsymbol{\Phi}_{is}$ $(s = 1, ..., p)$ and $\boldsymbol{\Lambda}_{ir}$ $(r = 0, ..., q)$ are, respectively, $k_i \times k_i$ and $k_i \times k_i*$ coefficient matrices of the lagged endogenous and the lagged weakly exogenous variables, and $\boldsymbol{\varepsilon}_{it}$ is an error term with covariance matrix $\boldsymbol{\Sigma}_i$.

The country i weakly exogenous (or foreign) variables are calculated as a weighted average of the endogenous variables in all the other economies:

$$\mathbf{x}^*_{it} = \sum_{j=1}^{N} \omega_{ij} x_{jt} \tag{2.2}$$

with ω_{ij} representing a non-negative weight connecting country i and j. We assume that $\omega_{ii} = 0$ and $\sum_{j=1}^{N} \omega_{ij} = 1$. They reflect the relative magnitude of the interaction among economies and are usually approximated using data on bilateral trade flows. The underlying assumption of weak exogeneity of the foreign variables implies that most countries are small relative to the world economy.

By stacking the N country-specific models, we obtain the global one as an approximation of the large dimensional VAR :

$$\mathbf{G}_0 \mathbf{x}_t = \boldsymbol{\alpha}_0 + \sum_{q=1}^{Q} \mathbf{G}_q \mathbf{x}_{t-q} + \boldsymbol{\varepsilon}_t, \tag{2.3}$$

where $\boldsymbol{\alpha}_0$ is a $k \times 1$ vector of intercepts and $\mathbf{G}_q$ for $q = 0, ..., Q$ are $k \times k$ coefficient matrices embodying contemporaneous and lagged dependence among countries, with $k = \sum_i k_i$ and $Q = \max(p_1, p_2)$. Finally, $\boldsymbol{\varepsilon}_t$ is the global error term with block diagonal covariance matrix $\boldsymbol{\Sigma}$.

If the matrix $\mathbf{G}_0$ is invertible then, pre-multiplying equation (2.3) by $\mathbf{G}_0^{-1}$, we obtain the GVAR representation

$$\mathbf{x}_t = \mathbf{F}_0 + \sum_{q=1}^{Q} \mathbf{F}_q \mathbf{x}_{t-q} + \mathbf{G}_0^{-1} \boldsymbol{\varepsilon}_t, \tag{2.4}$$

with $\mathbf{F}_0 = \mathbf{G}_0^{-1} \boldsymbol{\alpha}_0$ and $\mathbf{F}_q = \mathbf{G}_0^{-1} \mathbf{G}_q$ for $q = 1, ..., Q$.

Bayesian estimation

The standard GVAR literature assumes that $\mathbf{x}_{it}^*$ are weakly exogenous, and therefore OLS is the most widely used estimation technique. However, such assumption can be violated when large economies are included in the analyzed sample. For this reason, it is preferable to perform maximum likelihood estimation of the GVAR (Elhorst et al., 2021).

In order to cope with the possible presence of dominant units, we refer to the maximum likelihood estimator as a special case of a Bayesian GVAR. Indeed, at the same time such an approach: a) allows the researcher to specify her/his prior beliefs on the parameters of interest; b) induces country-specific degrees of shrinkage on the parameters, which improves forecast in a significant way.

The prior implementation is facilitated by rewriting each country model compactly as:

$$\mathbf{x}_{it} = \mathbf{\Pi}_i^{'}\mathbf{z}_{it-1} + \boldsymbol{\varepsilon}_{it}, \tag{2.5}$$

where $\mathbf{z}_{it-1} = [1, \mathbf{x}_{it-1}^{'}, ..., \mathbf{x}_{it-p_1}^{'}, \mathbf{x}_{it}^{*'}, ..., \mathbf{x}_{it-p_2}^{*'}]^{'}$ is the K_i dimensional vector of regressors, with $K_i = 1+k_ip_1+k_i^*(p_2+1)$, and $\mathbf{\Pi}_i^{'} = [\boldsymbol{\alpha}_{i0}, \mathbf{\Phi}_{i1}, ..., \mathbf{\Phi}_{ip_1}, \mathbf{\Lambda}_{i0}, ..., \mathbf{\Lambda}_{ip_2}]^{'}$ is the associated matrix of stacked coefficients. The model specification is completed by assuming that $\boldsymbol{\varepsilon}_{it}$ are normally distributed:

$$\boldsymbol{\varepsilon}_{it} \sim \mathcal{N}(0, \mathbf{\Sigma}_i). \tag{2.6}$$

In order to derive a compact form, equation (2.5) can be rewritten as a multivariate linear regression:

$$\mathbf{X}_i = \mathbf{Z}_i\mathbf{\Pi}_i + \boldsymbol{\varepsilon}_i, \tag{2.7}$$

where $\mathbf{X}_i$ is a $T \times k_i$ matrix of endogenous variables, $\mathbf{Z}_i$ is a $T \times K_i$ matrix of stacked explanatory variables and $\boldsymbol{\varepsilon}_i$ is a matrix of errors.

Following Cuaresma et al. (2016), we employ the Normal-Inverse Wishart prior, which belongs to the class of natural conjugate priors. Let $\mathbf{\Psi}_i = \text{vec}(\mathbf{\Pi}_i)$ denote the $v_i = K_ik_i$ dimensional coefficient vector. The prior setting reads as:

$$\mathbf{\Psi}_i|\mathbf{\Sigma}_i \sim \mathcal{N}(\underline{\mathbf{\Psi}}_i, \mathbf{\Sigma}_i \otimes \underline{\boldsymbol{V}}_i), \tag{2.8}$$

$$\mathbf{\Sigma}_i \sim \mathcal{IW}(\underline{S}_i, v_i). \tag{2.9}$$

Following the literature on Bayesian VARs (Litterman, 1986; Sims and Zha, 1998), we assume that the variables in the system follow simple random

walks. This results in imposing $\underline{\Psi}_i = 1$ for the entries concerning the first own lag of each endogenous variable, $\underline{\Psi}_i = 0$ otherwise. Regarding the elicitation of $\mathbf{V}_{i0}$, we specify a Minnesota prior. In particular, for the parameters in $\mathbf{\Psi}_i$ corresponding to lag r of variable g, it is given by:

$$\underline{\boldsymbol{V}}_{i,gr} = \begin{cases} \frac{\alpha_1}{r^k \mathbf{\Sigma}_{ig}} & \text{for the } r-th \text{ lag of variable } g \\ \frac{\alpha_2}{(1+r)^k \mathbf{\Sigma}^*_{ig}} & \text{for the } r-th \text{ lag of variable } g \text{ if weakly exogenous} \\ \alpha_3 & \text{for the constant term} \end{cases} \tag{2.10}$$

where α_1 and α_2 are hyperparameters controlling the tightness of the prior on the endogenous and weakly exogenous parts respectively, $\mathbf{\Sigma}_{ig}$ and $\mathbf{\Sigma}^*_{ig}$ are standard deviations obtained by estimating univariate regressions on the g-th endogenous variable and g-th exogenous one, α_3 controls the tightness of the prior on the constant term. For a detailed explanation of the prior setup see Cuaresma et al. (2016).

Given the high cross-sectional dimension handled by the GVAR, it is not necessary to employ a long lag length as in a standard VAR (Burriel and Galesi, 2018). Hence we set $p_1 = p_2 = 1$ for the country models. Regarding the hyperparameters, we remain uninformative through relatively flat priors both for the mean and variance parameters. Even though the total cross-sectional dimension in the GVAR is large, this econometric technique naturally acts as a dimensionality method. In our study, we employ a relatively small number of countries and variables, thereby there is no real necessity to regularize mean and variance parameters through tight priors. As a direct consequence, we opt for uninformative priors by setting $\alpha_1 = \alpha_2 = \alpha_3 = 100$ for $\mathbf{\Psi}_i$, and $\underline{S}_i = \mathbf{I}_{K_i k_i}$, $v_i = k_i + 2$. This prior degree of freedom is the minimum level such that the expected value of $\underline{S}_i$ is defined.

Forecast error variance decomposition in Global VARs

Our approach stems from the econometric connectedness measures developed by Diebold and Yilmaz (2009, 2012, 2014), which generally assume a stationary approximating time series VAR model, and that we apply to the GVAR modeling framework. Diebold and Yilmaz (2014) build generalized forecast error variance decomposition (GFEVD) on approximating Vector Autoregressive (VAR) models. The idea behind this approach is that the more a variable is important in forecasting the future dynamics of another, the more a shock in the former impacts the time series trajectory of the latter.

More recently, in line with our methodology, Greenwood-Nimmo et al. (2021) develop a method to employ forecast error variance decompositions to evaluate the macroeconomic connectedness in any multi-country macroeconomic model with an approximate VAR representation. They build their research on Diebold and Yilmaz (2014) and, particularly, on the shortcoming of not correctly dealing with high-dimensional cross-section data series and with an increasing number of variables m. This framework presents several advantages. In particular, it is convenient in terms of interpretability, it is able to take into account for contemporaneous impacts and it can predictively measure system-wide and pairwise connectedness in terms of direction and magnitude of network links.

The methodology by Greenwood-Nimmo et al. (2021) overcomes both the curse of dimensionality, which generally calls for a low number of variables m in empirical applications, and for the need of accommodating intermediate levels of aggregation in the modelling strategies, in a multilevel perspective. They propose a simple approach to overcome both issues based on re-normalisation and block aggregation of the connectedness matrix. The exposition of the block aggregation routine exploits the fact that GFEVDs are invariant to the ordering of variables in the VAR model.

Given the nature of the variables underlying the Global VAR dynamical

system, we follow Greenwood-Nimmo et al. (2021) who propose an aggregation scheme for the GFEVD in order to reduce its dimensionality with a direct interpretation on the countries (variables) FEVD. The GFEVD assumes the form:

$$\underset{(m\times m)}{\mathbb{C}_R^{(h)}} = \begin{bmatrix} \phi_{1\leftarrow 1}^{(h)} & \phi_{1\leftarrow 2}^{(h)} & \cdots & \phi_{1\leftarrow k}^{(h)} \\ \phi_{2\leftarrow 1}^{(h)} & \phi_{2\leftarrow 2}^{(h)} & \cdots & \phi_{2\leftarrow k}^{(h)} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_{k\leftarrow 1}^{(h)} & \phi_{k\leftarrow 2}^{(h)} & \cdots & \phi_{k\leftarrow k}^{(h)} \end{bmatrix}, \tag{2.11}$$

where $\phi_{i\leftarrow j}^{(h)}$ represents the contribution of variable j to the h -steps-ahead forecast error variance of variable i. Similarly, $\phi_{i\leftarrow i}^{(h)}$ denotes the contribution of variable i to its own h-steps-ahead forecast error variance (see Appendix (A.1) for further details on the connectedness matrix).

The (i, j) -th element of $\mathrm{C}_R^{(h)}$ represents the proportion of the total h -steps-ahead FEV of the system accounted for by the spillover effect from variable i to variable i. With this modification we are ensured that we may achieve a percentage interpretation even after aggregating groups of variables in the system.

Suppose we are interested in analyzing the spillover measures developed by Diebold and Yilmaz (2014) but focusing on the countries. It makes sense to aggregate the FEVD according to country blocks, where each ith element is obtained as an aggregation (sum) of the ith country block with its own variables. If, instead, the aim is to carry a variable analysis, the aggregation can be done considering the variable blocks.

Once we have collected $\mathbf{x}_t$ into b groups $\mathbb{C}_R^{(h)}$ can be equivalently expressed as:

$$\underset{(m\times m)}{\mathbb{C}_R^{(h)}} = \begin{bmatrix} \mathbf{B}_{1\leftarrow 1}^{(h)} & \mathbf{B}_{1\leftarrow 2}^{(h)} & \cdots & \mathbf{B}_{1\leftarrow b}^{(h)} \\ \mathbf{B}_{2\leftarrow 1}^{(h)} & \mathbf{B}_{2\leftarrow 2}^{(h)} & \cdots & \mathbf{B}_{2\leftarrow b}^{(h)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{b\leftarrow 1}^{(h)} & \mathbf{B}_{b\leftarrow 2}^{(h)} & \cdots & \mathbf{B}_{b\leftarrow b}^{(h)} \end{bmatrix}. \tag{2.12}$$

Consider now all the blocks lying on the main diagonal of that matrix (i.e. $\mathbf{B}_{k\leftarrow k}^{(h)}$); they contain all of the within-group FEV contributions. We can therefore define the within-group FEV contribution for the k th group as follows

$$\mathcal{W}_{k\leftarrow k}^{(h)} = \mathbf{e}_{m_k}' \mathbf{B}_{k\leftarrow k}^{(h)} \mathbf{e}_{m_k}, \tag{2.13}$$

where $\mathbf{e}_{m_k}$ is the usual selection vector. Roughly speaking, the within-group FEV contribution for the k th group is equal to the sum of the elements of the block $\mathbf{B}_{k\leftarrow k'}^{(h)}$. Analogously, $\mathbf{B}_{k\leftarrow \ell}$ for $k \neq \ell$ relates to the transmission of information across groups. Hence we define the spillover from group ℓ to group k as $\mathcal{F}_{k\leftarrow \ell}^{(h)} = \mathbf{e}_{m_k}' \mathbf{B}_{k\leftarrow \ell}^{(h)} \mathbf{e}_{m_\ell}$ and the spillover to group k from group ℓ as $\mathcal{T}_{\ell\leftarrow k}^{(h)} = \mathbf{e}_{m_\ell}' \mathbf{B}_{\ell\leftarrow k}^{(h)} \mathbf{e}_{m_k}$. By collecting all these measures, we can define the h-step ahead block connectedness matrix of dimension $b \times b$ as

$$\underset{(b\times b)}{\mathbb{B}^{(h)}} = \begin{bmatrix} \mathcal{W}_{1\leftarrow -1}^{(h)} & \mathcal{F}_{1\leftarrow -2}^{(h)} & \cdots & \mathcal{F}_{1\leftarrow b}^{(h)} \\ \mathcal{F}_{2\leftarrow 1}^{(h)} & \mathcal{W}_{2\leftarrow 2}^{(h)} & \cdots & \mathcal{F}_{2\leftarrow b}^{(h)^{(h)}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{F}_{b\leftarrow 1}^{(h)} & \mathcal{F}_{b\leftarrow 2}^{(h)} & \cdots & \mathcal{W}_{b\leftarrow b}^{(h)} \end{bmatrix}. \tag{2.14}$$

Note that the dimension of this grouped matrix is $b^2 < K^2$, which implies a significant improvement on the FEVD interpretation in large models ease the processing constraints encountered in large models. It is now straightforward to develop aggregate connectedness measures at the group level.

Within this framework, the total *from*, *to* and *net* connectedness of the kth group are respectively defined as follows:

$$\mathcal{F}_{k\leftarrow\bullet}^{(n)} = \sum_{\ell=1,\ell\neq k}^{b} \mathcal{F}_{k\leftarrow\ell}^{(n)}, \tag{2.15}$$

$$\mathcal{T}_{\bullet\leftarrow k}^{(n)} = \sum_{\ell=1,\ell\neq k}^{b} \mathcal{T}_{\ell\leftarrow k}^{(n)}, \tag{2.16}$$

$$\mathcal{N}_{\bullet\leftarrow k}^{(n)} = \mathcal{T}_{\bullet\leftarrow k}^{(n)} - \mathcal{F}_{k\leftarrow\bullet}^{(h)}, \tag{2.17}$$

where $\mathcal{F}_{k\leftarrow\bullet}^{(\mathrm{h})}$ measures the total spillover from all other groups to group k, the total *from* contribution affecting group k. $\mathcal{T}_{\bullet\leftarrow k}$ measures the total spillover to all other groups from group k, the total *to* contribution arising from group k. $\mathcal{N}_{\bullet\leftarrow k}^{(n)}$ is the net connectedness of group k. The subscript $i \leftarrow \bullet$ indicates that the directional effect is from all other variables to variable i. It is possible to derive in a similar way the aggregate heatwave and spillover indices, expressed in terms of the b groups, as:

$$\mathcal{H}^{(n)} = \sum_{k=1}^{D} \mathcal{W}_{k\leftrightarrow k}^{(h)}, \tag{2.18}$$

$$\mathcal{S}^{(n)} = \sum_{k=1}^{b} \mathcal{F}_{k\leftarrow\bullet}^{(n)} \equiv \sum_{k=1}^{b} \mathcal{T}_{\bullet\leftarrow k}^{(n)}, \tag{2.19}$$

where $\mathcal{H}^{(n)} + S^{(h)} = 1$ and $\sum_{k=1}^{D} \mathcal{N}_{\bullet\leftarrow k}^{(h)} = 0, \forall h$ by construction. Differently from the variable-level heatwave and spillover measures, in $\mathcal{H}^{(\mathrm{h})}$ and $\mathcal{S}^{(\mathrm{h})}$ measure the heatwave and spillover effects in a consistent way with respect to an aggregation routine.

From above, one can derive also two indices which are key to evaluate interconnectedness among entities and small group of entities, such as geopolitical units and financial market sectors. They express either how dependent is the $k-th$ group on external conditions or to which degree does the $k-th$

group influences the $k-th$ group or it is influenced by the system as a whole. The first measure is the dependence index:

$$\mathcal{O}_k^{(h)} = \frac{\mathcal{F}_{k\leftarrow\bullet}^{(h)}}{\mathcal{W}_{k\leftarrow k}^{(h)} + \mathcal{F}_{k\leftarrow\bullet}^{(h)}}. \tag{2.20}$$

The index measures the relative importance of external shocks for the $k-th$ group. In particular, as $\mathcal{O}_k^{(\mathrm{h})} \to 1$, the network structure of group k is dominated by external shocks, while group k is unaffected by external shocks if $\mathcal{O}_k^{(h)} \to 0$. Similarly, the influence index can be derived as:

$$\mathcal{I}_k^{(h)} = \frac{\mathcal{N}_{\bullet\leftarrow k}^{(h)}}{\mathcal{T}_{\bullet\leftarrow k}^{(h)} + \mathcal{F}_{k\leftarrow\bullet}^{(h)}}, \tag{2.21}$$

where $-1 \leq \mathcal{I}_k^{(h)} \leq 1$. For any horizon h, the $k-th$ group is a net shock recipient if $-1 \leq \mathcal{I}_k^{(h)} < 0$, a net shock transmitter if $0 < \mathcal{I}_k^{(h)} \leq 1$, and neither of the two if $\mathcal{I}_k^{(h)} = 0$. Hence, the influence index expresses the extent to which the $k-th$ group influences or is influenced by conditions in the system. When studying connectedness among countries, the coordinate pair $\left(\mathcal{O}_k^{(h)}, \mathcal{I}_k^{(h)}\right)$ in dependence-influence space can give a good representation of country (or variable) k and its role in the global network system. While small open economies would tend to be located close to the coordinates $(1, -1)$, a dominant economy would lay in the proximity of $(0, 1)$.

We remark that the GFEVD has the inherent characteristic of not reflecting the impact of structural shocks, but only composite ones, an approach which suffers from the contemporaneous correlation across equation innovations. The GFEVD, as opposed to the orthogonal FEVD, comes without the necessity of devising an identification schemes which could result into questionable choices.

Network theory and forecast error variance decompositions

In general, a variety of network analysis approaches, which study the structure and interactions of weighted directed network, can be applied to any econometric connectedness matrix derived from GFEVD. Another important issue is the dimensionality of the Global GFEVD directional spillover network matrix. Indeed, given the ability of the GVAR to handle large number of variables, it allows to derive high-dimensional GFEVD matrices. However, these matrices are fully connected. This calls for the implementation of tools borrowed from network theory and topological analysis to study backbone structures, centrality measures, hub-authority relationships and many more network analysis approaches on econometric connectedness measures.

In this context, we propose to merge two very consolidated approaches - i.e. financial econometrics and network theory - to investigate dominant links existing in the matrix of time series GFEVD from two different perspectives, and, at the same time, in a complementary way. In particular, we investigate the network topology of the risk spillover matrix coming from the Global VAR generalized forecast error variance decomposition. In other words, we employ the MST, the Louvain community detection and the Kleinber centrality measure so to study the backbone structure and the hub-authority dynamics of macroeconomic relationships, on indices based on the decomposition of the variance-covariance matrix of the forecast error derived from the dynamic GVAR estimates.

The combination of the two methodologies presents two main practical advantages. First, we are able to statistically capture systemic risk in a multi-country and multi-variate way, either at the aggregate or individual level, and within a predictive framework. Second, we benefit from enhanced explainability of results and insights derived from the network analysis of econometric measures of systemic risk in the two specific dimensions of the

longitudinal time series.

Minimum Spanning Tree

In order to simplify the relationships given by the GFEVD matrix, we apply the Minimum Spanning Tree (MST) representation. This is a widely employed technique in finance - see, for instance, Miccichè et al. (2003); Musciotto et al. (2018). The Kruskal's algorithm (Kleinberg, 1998), which delivers the minimum spanning tree from a given adjacency matrix, is able to reduce the number of links among the nodes by drawing a link for each node to its closest neighbour. In this way, we achieve dimensionality reduction of the GFEVD spillover matrices obtained from the Global VAR estimation, and we are able to determine the backbone structure of both countries and variables.

The methodological premise of the Minimum Spanning Tree (MST) is the existence of a given weighted graph $G = [V, E]$, with $v \in V$ vertices and $e \in E$ weighted edges. Without loss of generality, the weights can be thought as the cost to reach v_j from v_i. We are interested in finding a subgraph $\tilde{G}$ of G such that $V(G) = V(\tilde{G})$ (i.e. they span the same nodes) but with the minimum possible aggregate cost, thought as the sum of all the graph weights.

Let us define $ST(G)$ as the spanning tree set generated by the graph G, i.e. the set of all trees that spans all the nodes of G. $\hat{G} \in ST(G)$ is called a minimum weight spanning tree if the sum of the weights of the edges of $\hat{G}$ does not exceed the sum for any other spanning tree of G (i.e. the one that minimizes the sum of weights of the set $ST(G)$), see Harris et al. (2008).

There are a variety of algorithms for exploiting minimum weight spanning trees. The most widespread and employed is the Kruskal's algorithm, which can be summarized as follows. Given a connected, weighted graph G:

1. Find an edge of minimum weight and mark it.

2. Among all of the unmarked edges that do not form a cycle with any of the marked edges, choose an edge of minimum weight and mark it.

3. If the set of marked edges forms a spanning tree of G, then stop. If not, repeat step 2.

Louvain clustering for Community detection

An important feature of complex networks is its community structure, which refers to the presence of groups of nodes. Each group contains a set of nodes; within each group the density of edges is higher than that among the groups. Community detection algorithms might reveal the hidden relations among the nodes in a network. The core of the Louvain method (Blondel et al., 2008) is to find a partition of the vertex set that maximizes the modularity of the considered graph. This function provides a way to value the existence of an edge between two vertices of an undirected network by comparing it with the probability of having such an edge in a random model.

Modularity is a measure of the topology of networks, representing the degree to which a network is fragmented into communities. A high modularity corresponds to a network with dense links within community nodes, but sparse links between the nodes belonging to different communities. This is of relevance when analyzing financial and economic spillover networks, since it reveals information on the communities involved in the risk propagation mechanism, as well as on shock-resilient entities. Formally, the modularity Q of a partition $\mathcal{C}$ of an undirected graph $G = (V, E)$ is defined as follows :

$$Q = \frac{1}{2m} \sum_{i,j} \left[A_{ij} - \frac{d_i d_j}{2m} \right] \delta\left(c_i, c_j\right) \tag{2.22}$$

where m stands for the number of edges of G, A_{ij} represents the weight of the edge between i and j, d_i is the degree of vertex i (i.e. the number of

neighbors of i), c_i is the community the vertex i belongs to and δ is the so called Kronecker delta function, defined as $\delta(u, v) = 1$ if $u = v$, 0 otherwise.

Louvain's algorithm relies on a greedy procedure. First, each node is assigned to its own community. Then the algorithm tries to increase the value of modularity by moving nodes into the community of each neighbor. In other words, the algorithm computes the gain of modularity obtained by adding vertex i to community C as follows:

$$\Delta_Q = \left[\frac{\sum_{in} + d_i^C}{2m} - \left(\frac{\sum_{tot} + d_i}{2m}\right)^2\right] - \left[\frac{\sum_{in}}{2m} - \left(\frac{\sum_{tot}}{2m}\right)^2 - \left(\frac{d_i}{2m}\right)^2\right] \tag{2.23}$$

where d_i^C denotes the degree of node i in community C, $\sum_{in}$ the number of edges contained in community C and $\sum_{\text{tot}}$ the total number of edges incident to community C.

Once this value is derived for all communities i is connected to, i is placed where the value of the modularity is maximized. If no increase is possible, i remains in its original community. This process is applied repeatedly and sequentially to all nodes until there are no moves that improve the value of modularity. We remark that the main difference between the Louvain and K-means methodologies is that the former is nonparametric, and requires no a priori assumptions on the graph. Moreover, K-means and most of the clustering techniques act on data points embedded in a space, while Louvain on data points linked by an underlying graph.

Hubs and authorities

In this Chapter we employ a well-known centrality measure in order to explore the centrality and, thereby, hubs, authorities and systemically important nodes in the Global VAR macroeconomic system GFEVD framework. Network theory includes several centrality measures such as the degree centrality, counting how many neighbours a node has, as well centrality measures

based on properties of graphs. Among centrality measures we remark Katz's centrality - see (Katz, 1953) - PageRank - (Brin and Page, 1998), hub and authority centralities - (Kleinberg, 1999) - and the eigenvector centrality - (Bonacich, 2007).

In our context, we aim at investigating hub-authority network centralities of country and variable systems, which can be derived through the Kleinber centrality measure. According to such centrality measure, a node is important if it has many incoming links from other important nodes. In general, nodes with no incoming links cumulate, in the best case, only to a minimum amount of centrality. However, it can also be argued that a node is important, even if not pointed by many others, if it links to a set of important vertices. Hence, there are two kinds of central nodes: *authorities*, that contain reliable information, and *hubs*, that suggest (point) where to find reliable information. The first ones are pointed by many good hubs nodes, and, conversely, the second ones point to many good authorities vertices.

For each node of a graph, we use $au(x)$ to denote its authority score and $hu(x)$ its hub score. We start by setting $au(x) = hu(x) = 1$ for all nodes. The core of the algorithm is an iterated update of the hub and authority scores of all nodes given by equation (2.24), which capture the intuitive notions that good hubs point to good authorities and that good authorities are pointed to by good hubs:

$$\begin{aligned} hu(x) &= \sum_{k \in Pa(x)} au(k), \\ au(x) &= \sum_{k \in Ch(x)} hu(k), \end{aligned} \tag{2.24}$$

where $Pa(x)$ is the set of parents of node x and $Ch(x)$ is the set of children of x. The first line of equation (2.24) sets the hub score of a node v to the sum of the authority scores of the nodes it links to. In other words, if v links to nodes with high authority scores, its hub score increases. The second line plays the reverse role; if node v is linked to by good hubs, its authority score

increases. Let us shift equation (2.24) into matrix form. Let Au and Hu denote the vectors of all hub and all authority scores respectively. Let D denote the adjacency matrix of this graph: it is a square matrix with entry $D_{ij} = 1$ if there a edge from node i to j, and 0 otherwise. Then, we have:

$$\begin{aligned} Hu &= D'Au, \\ Au &= DHu. \end{aligned} \tag{2.25}$$

Substituting Au (Hu) in the second (first) equation with the right hand of the first (second) one we can rewrite equation (2.25) as:

$$\begin{aligned} Hu &= DD'Hu \\ Au &= D'DAu. \end{aligned} \tag{2.26}$$

If we introduce the (unknown) eigenvalue, the first line of equation (2.26) becomes the equation for the eigenvectors of DD', and the second becomes the equation for the eigenvectors of $D'D$:

$$\begin{aligned} Hu &= \frac{1}{\lambda_a} DD'Hu \\ Au &= \frac{1}{\lambda_h} D'DAu \end{aligned} \tag{2.27}$$

where λ_a and λ_h are, respectively, the eigenvalues of DD' and $D'D$. The iterative process introduced in equation 2.24 is equivalent to the problem of solving the characteristic equation of matrices by finding the principal eigenvectors (and eigenvalue) of DD' and $D'D$. The resulting computation thus takes the following form, once the weighted graph is available:

1. Compute $D'D$ and $D'D$
2. Compute the principal eigenvectors of $D'D$ and DD' to form the vector of hub scores Hu and authority scores Au.

3. Rank the vectors found to check which subset of nodes in the graph have the highest hub and authority values.

2.3 Data

Our approach combines real economy and economic sentiment indicators, as measured by the Economic Sentiment indices survey data provided by EUROSTAT. We analyze a selected set of real economy and survey sentiment data variables for 12 countries in Europe, namely: Austria, Belgium, Denmark, France, Germany, Ireland, Italy, Netherlands, Poland, Spain, Sweden and United Kingdom. Such choice is due to data availability and quality constraints thus, the analyzed 12 countries, assure full and complete time series. Nevertheless, these 12 countries cover more than 90% of the European Union's GDP [1]. As Great Britain left the EU in January 2020, we rely on previous years statistics. In particular, we analyze the time series of Industrial Production, Retail Trade and Economic Sentiment over the period 01/2000-11/2021. We select four crucial periods for the analysis of critical transitions in the real economy and economic sentiment:

- Phase 1: The pre-crisis period (01/2006-08/2008);
- Phase 2: The Global Financial and European Sovereign Debt crises period (09/2008-12/2012);
- Phase 3: the post-crisis period (01/2013-02/2020);
- Phase 4: The COVID-19 pandemic period (03/2020-11/2021).

Before the modelling exercise, we test the assumption that the country-specific foreign variables are weakly exogenous, i.e. we test for the presence of dominant units.

[1] Source: EUROSTAT Database.

This ensures the consistent estimation of the number of cointegrating relationships and the cointegrating vector for each country model in a separate way. We refer the reader to Chudik and Pesaran (2013); Konstantakis et al. (2015); Pesaran and Yang (2020) for thorough discussions on dominant units. Details and results on the weak exogeneity test are contained in A.1.

In summary, the weak exogeneity assumptions are violated in 3 cases out of 21. The first one pertains Industrial Production of UK, which is consistent with Dees et al. (2007), who found a rejection in the UK country model. The second and third ones are related to Retail Trade of Austria and Sweden. Given these premises, we have opted for a more consistent estimation procedure through the Bayesian GVAR outlined in the methodological section.

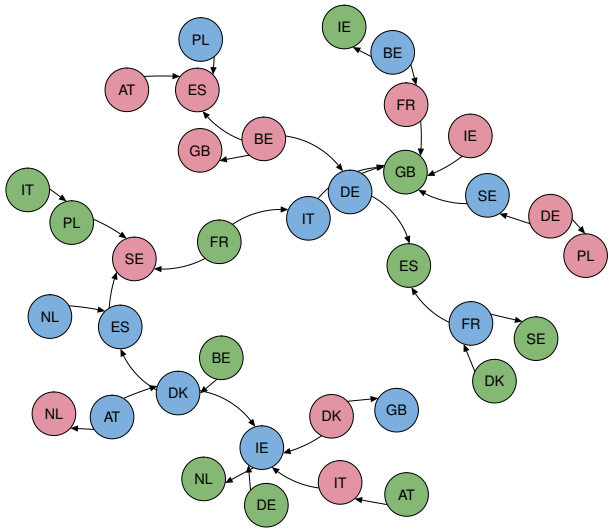
2.4 Empirical results

Global analysis

Our global analysis examines the interaction between real economy and sentiment both at the country and variable levels. In particular, we first show the degree of cointegrating relationships existing among our data variables. After that, we show the network structure of the GFEVD of the three variables of interest - i.e. Industrial Production, Retail Trade and Sentiment indices - across the 12 European countries under consideration.

Figure 1 shows a graphical representation of the Minimum Spanning Tree calculated on the generalized forecast error variance decomposition (GFEVD) relative to the whole sample period. Network nodes represent system variables, whereas edges represent the directional contribution of each variable to the forecast error variance decomposition of the others. The retained MST paths are represented by edges with larger weights of GFEVD transmitted to the other nodes, highlighting the backbone structure of the GFEVD spillover

Figure 1: Global forecast error variance MST



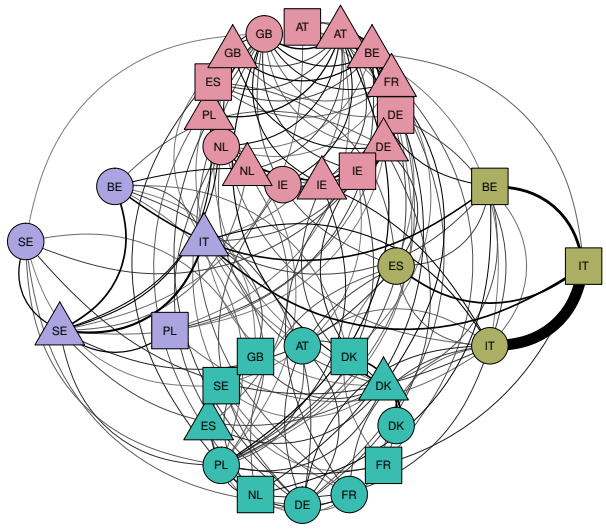
MST calculated on the whole sample GFEVD. Nodes represent system variables, whereas edges represent the directional contribution of each variable to the forecast error variance decomposition of the others. Colours stand for Industrial Production (red), Retail Trade (green) and Economic Sentiment (blue). Self-loops are omitted.

network.

We notice a central role of the French Retail Trade index, which acts as a bridge and shock transmitter towards two consistent sub-graphs, particularly by means of its link with the Industrial Production index of Estonia and with the Economic Sentiment of Italy. The two sub-graphs, divided in this way, differ significantly in composition. The first one is mainly composed by Industrial Production variables, with Austria, Spain, United Kingdom and Belgium being - directly or indirectly - linked.

The second sub-graph is mainly composed by Economic Sentiment and

Figure 2: Global forecast error variance clustering



Clustered GFEVD network structure of real economy and economic sentiment variables. Groups are classified according to the Louvain clustering algorithm. Squares: Industrial Production; triangles: Retail Trade; circles: Economic Sentiment.

Retail Trade indices. In this case, sentiment variables of Netherlands, Spain, Denmark, Austria and Ireland influence each other in the context of the network backbone structure.

Figure 2 shows the clustered GFEVD network structure of real economy and economic sentiment variables over the full sample period, where clusters are obtained through the Louvain community detection algorithm. Results highlight the emergence of four clusters: two of them are clearly denser than the other two. The largest cluster (pink) mainly consists of Retail Trade variables, whereas the second largest (blue) shows a prevalence of Economic Sentiment and Industrial Production variables, with many country-specific

connections. Interestingly, the GFEVD network clusters show the emergence of two relatively isolated, but still communicating vessels. The first one (purple) is composed by the Industrial Production of Poland, the Retail Trade of Italy and Sweden, along with the Economic Sentiment of Sweden and Belgium. In the second group (green), we find the Industrial Production indices of Italy and Belgium and the Economic Sentiment of Italy and Spain. The strongest link is that between the Economic Sentiment and the Industrial Production of the Italy, which has strongly co-moved over the last two decades, and acts as a main information bridge across clusters.

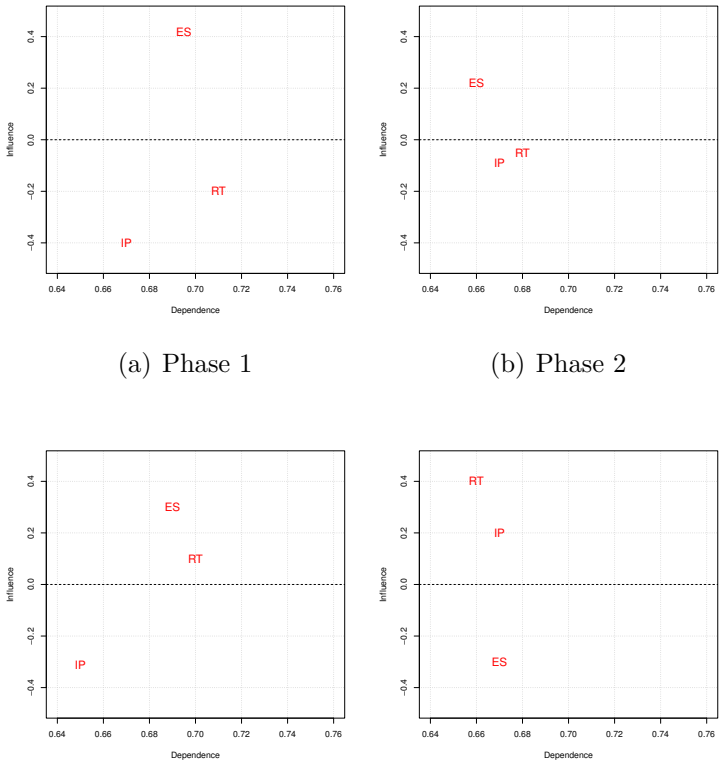
Variable analysis

The aim of the variable analysis is to investigate the informationally dominant variables, in terms of GFEVD transmitted to others, in the network of variables (Industrial Production, Retail and Sentiment indices) over time. The analysis at a variable level aggregates country GFEVD and derives overall predictive spillovers from one variable to another. Based on this, we derive the aggregate dependence and influence indices of the sentiment and real economic variables, so to measure the role played in the system at an aggregate level.

In this regard, Figure 3 shows the scatter plot obtained by generating a cartesian plane with the influence index on the x axis and the dependence index on the y axis, as defined in Section 2.2. Both the indices are calculated starting from the aggregation of the three variables of interest (Industrial Production, Retail and Sentiment indices) at the country level of the GFEVDs.

From the temporal dynamics along the four considered periods, we observe that the three first phases are quite similar, the main difference lies in the position of the Retail Trade index which swings between negative and positive values with regards to the influence index. That said, the most inter-

Figure 3: Dependence-influence variable relationships



(a) Phase 1

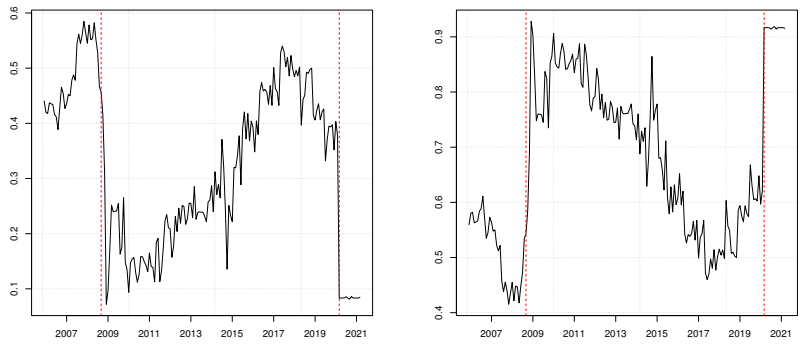
(b) Phase 2

(c) Phase 3

(d) Phase 4

Scatter plot obtained by generating a cartesian plane with the influence index on the x axis and the dependence index on the y axis. "IP" stands for Industrial Production, "RT" for Retail Trade and "ES" for Economic Sentiment.

Figure 4: Country Aggregate Heatwave and Spillover index



(a) Aggregate Heatwave (b) Spillover index

Aggregate Heatwave and Spillover index based on the country dynamic GFEVDs. The beginning of the Global Financial Crisis (September 2008) and the COVID-19 outbreak (February 2020) are marked in red.

esting pattern appears, once again, when considering the last period of crisis: the ranking of the indices is inverted and in particular the sentiment index presents a negative value for the influence, meaning that it is more influenced than before. This might be linked to the fact that, being COVID-19 a large and, most of all, unexpected exogenous shock to the network of real economic and sentiment interrelationships, it does cause a reversal of trend between sentiment and the real economy. While macroeconomic expectations have mostly tended to anticipate real macroeconomic dynamics in view of current situations, such a strong unexpected shock has swapped the relationship and led the real economic conditions to dominate the dynamics of sentiment.

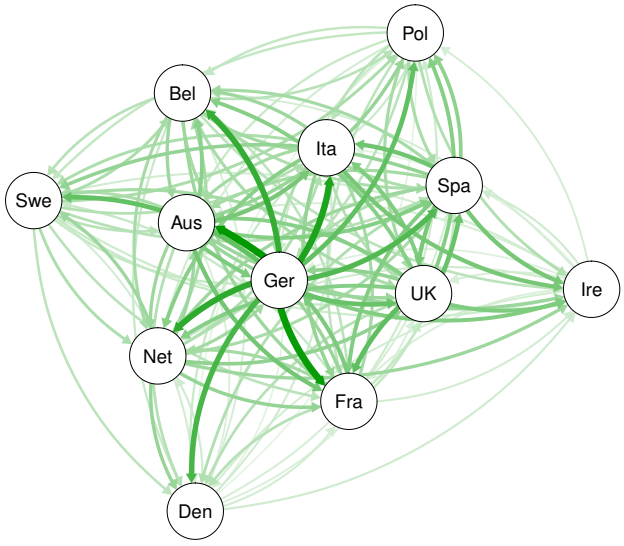
Country analysis

In the country analysis, we investigate country-level connectedness in terms of time series dynamics of real economic and sentiment variables. In particular, we first show the dynamic Aggregate Heatwave and Spillover index, so to give an idea of the magnitude of the interaction within- and cross-country GFEVD. Then, we examine the dependence and influence of each of the European countries in terms of transmitted/received economic and sentiment GFEVD spillovers. Finally, we study the hub-authority relationships across countries in terms of their contribution to the others with regards to the forecast error variance.

Figure 4 shows the Aggregate Heatwave and the Overall Spillover index based on the dynamic GFEVDs. The beginning of the Global Financial Crisis (September 2008) and the COVID-19 outbreak (February 2020) are marked in red. The two crisis periods emerge clearly from the curves dynamics with huge drops (hypes) followed by recovery periods. In particular, evidence shows that during both the Global Financial Crisis and the COVID-19 pandemic, the interconnectedness (Aggregate Heatwave) within country variables decreases sharply, likely due to large misalignment in the real economy and sentiment time series dynamics. Similarly, the between-country interconnectedness (Spillover index) starkly rises because of increased connection between countries' real economies and sentiment. Interestingly, the COVID-19 impact seems to be more persistent in time than that exerted by the Global Financial Crisis in terms of spillover magnitude.

Figure 5 shows the directed weighted GFEVD spillover network obtained from the country aggregation of the whole sample GFEVD. Nodes represent countries, whereas links represent the magnitude of pairwise directed forecast error variance transmitted to others. The central role is played by Germany which shows strong directed links with eight countries. Some other countries such as Poland, Ireland and Denmark seem to be overall more pheripheral in the spillover network, as well as mostly GFEVD receivers than transmitters.

Figure 5: Directed GFEVD country spillover network



Directed GFEVD country spillover network. Nodes represent countries, whereas links represent the magnitude of pairwise directed forecast error variance transmitted to others. Self-loops are omitted.

Such results need to be, however, analyzed from a dynamic perspective in order to capture the evolution of country interrelationships over time. For this purpose, Figure 28 shows the same directed weighted GFEVD spillover network obtained from the country aggregation of the GFEVD phase by phase.

In Tables 1 and 2, we report our results regarding the hubs and authorities algorithm performed on the GFEVD network. In particular, Table 1 shows countries ranked in decreasing order along with their "Hubs" value (in brackets) calculated on the directed network of country block GFEVDs divided into the four sub-samples under consideration. On the other hand, Table 2 reports countries ranked in decreasing order along with their "Authorities"

Table 1: Country-level network hubs

	1st	2nd	3rd	4th	5th
Phase 1	Net	Bel	UK	Pol	Ire
	(12.860)	(11.333)	(10.173)	(10.033)	(8.385)
Phase 2	Net	UK	Pol	Ger	Pol
	(24.101)	(12.968)	(8.958)	(8.024)	(7.862)
Phase 3	Ger	Spa	Pol	Net	Ita
	(12.957)	(10.826)	(10.594)	(10.579)	(9.697)
Phase 4	Ger	UK	Pol	Fra	Net
	(17.768)	(14.555)	(11.065)	(9.427)	(9.426)

The table reports countries ranked in decreasing order along with their normalized "Hubs" values (in brackets) calculated on the directed network of country block GFEVDs divided into the four sub-samples under consideration.

values.

The premise of the hubs and authorities analysis is that there are two types of central nodes: authorities, that contain reliable information on the forecast error variance of the others, and hubs, that point to authoritative information on such forecast error variance. Results show that there is a noticeable variability in the hub-authority relationships across European countries as regards to dominance in terms of sentiment and real economy spillovers.

Interestingly, we find that Netherlands exhibits both an authoritative and hub behaviour during the first two phases, while it then loses its relevance in the network in the aftermath of the Global Financial Crisis. At the same time, Germany's hub score increases, meaning it starts pointing to reliable sources of shock transmission, both during the tranquil period before COVID-19 and in the midst of the pandemic itself. In accord with the results on the dependence index, we additionally find that there is no predominant

Table 2: Country-level network authorities

	1st	2nd	3rd	4th	5th
Phase 1	Net	Bel	Pol	UK	Den
	(10.380)	(9.300)	(9.064)	(8.993)	(8.245)
Phase 2	Net	Spa	Ger	UK	Swe
	(9.793)	(8.954)	(8.685)	(8.578)	(8.278)
Phase 3	Ger	Spa	Net	Pol	Ita
	(9.504)	(9.023)	(8.930)	(8.893)	(8.531)
Phase 4	Spa	Ire	Fra	Net	UK
	(8.341)	(8.340)	(8.340)	(8.338)	(8.337)

The table reports countries ranked in decreasing order along with their normalized "Authorities" values (in brackets) calculated on the directed network of country block GFEVDs divided into the four sub-samples under consideration.

authority during the COVID-19 period, meaning that such large exogenous shock has affected the lead-lag relationships across countries, levelling out the authoritative behaviour of dominant countries in the network.

2.5 Conclusion

We propose a methodology which takes root from the statistical and econometric literature concerning generalized forecast error variance decomposition of multivariate time series models. We build on a dynamic spillover network framework derived from a Global VAR model, a suitable macroeconometric model to consider for simultaneous multi-level interdependencies. We then propose to exploit concepts from network theory to analyze the backbone structure, to perform community detection and to examine the hub-authority dynamics of longitudinal time series networks, both across

countries and variables. We exemplify our method with an empirical application to study the network topology of spillover indices across a set of European countries, macroeconomic variables and sentiment data over the period 01/2000-11/2021.

We demonstrate the usefulness of our method to analyze the network structure of shock propagation in longitudinal time series and, in particular: a) the shortest paths of contagion; b) the clusters of shock transmission; c) the role of nodes in the risk transmission channels. Our main results show how different crises exert starkly diverse impacts on real economy, economic sentiment and on the entire countries' macroeconomic network structure. In particular, while the impact of the Global Financial Crisis shows different consequences on the network interrelationships of macroeconomic variables if compared to those of the COVID-19 outbreak, the latter seems to be more persistent over time.

Despite the merit of our method in linking econometrics and network theory to produce advances in a cutting-edge research ground, there is room for methodological developments and empirical applications at the intersection of two fields. While econometric models are often robust, yet not exploited from a network perspective, network models can constitute a precious contribution to the understanding of time series model outputs and, more generally, to the establishment of new cross-cutting disciplines. This is a promising avenue for future research not only in the context of real economy networks, but also on financial, social and any kind of networks.

3 Bayesian Stochastic Search for Matrix Autoregressive Models

Based on the paper: Celani, A., Jones, G., and Pagnottoni, P. (2022), "Bayesian Variable Selection for Matrix Autoregressive Models". *Available at SSRN*.

3.1 Introduction

The emergence of a massive amount of high-dimensional time series observed in matrix form gives birth to new modeling challenges in economics, finance, and related fields. The existing approaches to dimension reduction in high-dimensional multivariate time series analysis can be organized in two major classes: a) factor models (Bai and Ng, 2002; Forni et al., 2005; Lam et al., 2011); b) modeling with frequentist regularization or Bayesian methods (Rothman et al., 2010; Song and Bickel, 2011; Kock and Callot, 2015; Park and Casella, 2008; Bańbura et al., 2010; Gefang, 2014; Ahelegbey et al., 2016; Korobilis, 2021).

While most of the extant modelling paradigms are designed to encourage parsimony by treating observations as time series vectors, the collection of data at the intersection of two or more classifications naturally constitutes matrix or tensor-valued time series. For instance, the collection of panel data forms matrix-valued observations, whose rows might represent indicators and columns countries. In this context, univariate time series analysis focuses on one element of the matrix at a time. Vector and panel time series analysis deal with the co-movement of one row in the matrix. Modeling each dimension separately annihilates the multidimensional structure of data, and can therefore lead to a significant loss of information or efficiency.

A strand of literature has therefore developed estimation procedures and investigated theoretical and probabilistic properties of multidimensional time

series models, including matrix-valued ones (Hoff, 2015, 2011; Chen et al., 2021; Wang et al., 2019; Billio et al., 2022). However, when the matrix observation has large dimensions, the matrix autoregressive (MAR) model involves a large number of parameters, which requires further dimension reduction techniques in order to produce accurate estimation. So far, this has been primarily tackled through the introduction of factor autoregressive models for multidimensional time series (Wang et al., 2019; Chen and Fan, 2021; Gao and Tsay, 2021; Chen et al., 2022).

In parallel, the literature on high-dimensional statistics has widely studied sparse recovery in the context of normal linear regression models, particularly stemming from the seminal papers of George and McCulloch (1993, 1997) on variable selection via Gibbs sampling, hereafter identified as stochastic search variable selection (SSVS). More recently, Ročková and George (2014) propose a deterministic alternative to stochastic search based on an EM variable selection (EMVS) algorithm to quickly find maximum a posteriori probability (MAP) estimates. As in SSVS, the EMVS method combines a spike-and-slab regularization procedure for the discovery of active predictor sets with subsequent evaluation of posterior model probabilities, with the difference rendering an appealing alternative in terms of computational efficiency.

Fully Bayesian variable selection has recently attracted attention in the multivariate time series context, starting from George et al. (2008) who propose a stochastic search approach to selecting restrictions for vector autoregressive (VAR) models. Despite that, the EMVS framework and its convenient features have yet to be exploited in multivariate time series. Moreover, variable selection has only been considered in vector-valued time series models to date, highlighting the need of identifying restrictions in multidimensional autoregressive models. The absence of restrictions on the regression coefficients results in a potentially large number of parameters relative to the available data. With a limited number of observations over-parameterization might affect the precision of inference.

We propose a novel matrix autoregressive model where sparsity is induced both in a fully Bayesian and a scalable EM framework, which enables fast posterior mode identification. By deriving a compact form for MAR models borrowed from the tensor linear regression framework, we design a traditional estimation strategy using MCMC and a Bayesian EMVS procedure for fast posterior mode identification. The latter allows to significantly shrink the computational time required for the MCMC procedure, in line with the need of current estimation algorithms to deal with large-scale multidimensional time series.

We demonstrate the properties of our proposed model through simulations and examples in the application to macroeconomic data. Our simulation experiments show: a) the gain in small sample efficiency of the proposed estimators relative to maximum likelihood (ML) in high-dimensional sparse settings; and b) that the proposed estimators perform generally better than standard VARs and several competing alternatives suited for longitudinal data, while limiting computational intensity in the EMVS formulation. The empirical application to macroeconomic data confirms that the model is able to: a) handle high-dimensional longitudinal data; b) outperform competing alternatives in high-dimensional settings; and c) yield enhanced interpretability given by the autoregressive model in matrix form.

Our model can be readily extended to the tensor autoregressive (TAR) framework. In contrast to other work (Billio et al., 2022), our formulation encompasses a Tucker structure in the matrix coefficient rather than a PARAFAC decomposition. The latter can be viewed as a special (constrained) case of Tucker decomposition, with a super-diagonal core tensor and the same number of factor components for each dimension. While the PARAFAC decomposition has the advantages of being highly parsimonious and unique, it requires the autoregressive coefficient matrices to be of rank one, which might be extremely restrictive in many economic and financial applications. Conversely, the Tucker structure admits an arbitrary number of factor components in each mode, hence allowing to model dimension asym-

metric tensors and benefit from enhanced interpretability of mode-specific interrelationships.

The remainder proceeds as follows. In Section 3.2 we outline the proposed sparse MAR model; then we illustrate the MCMC and EMVS estimation procedures. Section 3.4 evaluates the model performance and computational times, compared to some key competitors, through a simulation study. In Section 3.5 we conduct an empirical application to macroeconomic data. Section 3.6 contains some concluding remarks.

3.2 Model and Prior Structure

The Matrix Autoregressive Model

The MAR model takes advantage of the original structure of the data by modeling matrix autoregressive dynamics in a bilinear form. This is of paramount importance, for instance, in panel data applications, where each observation at time t can be conceived as a matrix, whose rows represent indicators and columns represent countries.

For $n = 1, ..., N$, let $\mathbf{y}_{n,t} \in \mathbb{R}^{G\times 1}$ and set $\mathbf{Y}_t = [\mathbf{y}_{1,t}, ..., \mathbf{y}_{N,t}] \in \mathbb{R}^{G\times N}$. The conditional mean of the matrix observation at time t is expressed as the product of P lagged observations by two (left and right) autoregressive coefficient matrices, $\mathbf{A}_i \in \mathbb{R}^{G\times G}$ and $\mathbf{B}_i \in \mathbb{R}^{N\times N}$ for $i = 1, ..., P$, controlling the row and column effects respectively. The model takes the form

$$\begin{aligned} \mathbf{Y}_t &= \sum_{i=1}^{P} \mathbf{A}_i \mathbf{Y}_{t-i} \mathbf{B}_i' + \mathbf{E}_t, \\ \mathbf{E}_t &\sim \mathcal{MN}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2), \end{aligned} \tag{3.1}$$

where $\mathbf{E}_t \in \mathbb{R}^{G\times N}$ is a matrix white noise of the same dimension of $\mathbf{Y}_t$ with two symmetric positive definite covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$. We use $\mathcal{MN}$

to denote the multilinear normal distribution (Gupta and D.K., 1999; Ohlson et al., 2013). For a more detailed explanation of the model interpretation see Chen et al. (2021).

Let vec$(\cdot)$ be the usual vectorization of a matrix by stacking its columns and $\otimes$ the usual Kronecker product. Let $\mathbf{y}_{t-i} = \text{vec}(\mathbf{Y}_{t-i})$ for $i = 0, ..., P$ and $\mathbf{e}_t = \text{vec}(\mathbf{E}_t)$. The vectorized form of the MAR model in equation (3.1) is

$$\begin{aligned} \mathbf{y}_t &= \sum_{i=1}^{P} (\mathbf{B}_i \otimes \mathbf{A}_i)\mathbf{y}_{t-i} + \mathbf{e}_t, \\ \mathbf{e}_t &\sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1). \end{aligned} \tag{3.2}$$

The representation in equation (3.2) shows that the MAR(P) model can be regarded as a special case of a VAR(P) model, with an autoregressive coefficient matrix given by a Kronecker product of the two mode-specific matrices.

The model can be made more parsimonious, if more restrictive, by assuming that the row and column effects matrices are time invariant ($\mathbf{A}_i = \mathbf{A}$, $\mathbf{B}_i = \mathbf{B}$), while embedding the effects of lagged observations in a row vector $\mathbf{c} \in \mathbb{R}^{1\times P}$. The conditional mean of $\mathbf{Y}_t$ is a linear combination of its P lagged values, pre and post multiplied by $\mathbf{A}$ and $\mathbf{B}$ as in a MAR(1). If $c_i \in \mathbf{c}$, the model is then

$$\mathbf{Y}_t = \mathbf{A}\left(\sum_{i=1}^{P} c_i \mathbf{Y}_{t-i}\right)\mathbf{B}^{'} + \mathbf{E}_t \tag{3.3}$$

Although it can be expressed as a MAR, the model arises naturally as a special case of tensor autoregression (TAR), as will be shown in the sequel. We therefore denote the model in equation (3.3), MAR$^*(P)$.

An unrestricted VAR(P) estimates $(GN)^2P \in \mathcal{O}(n^5)$ parameters, but due to the Kronecker structure imposed on its coefficient matrices, the MAR(P)

model estimates only $(G^2 + N^2)P \in \mathcal{O}(n^3)$ parameters, where $\mathcal{O}(\cdot)$ denotes the order of parameter complexity. The further restriction imposed that yields the MAR$^*(P)$ results in $G^2 + N^2 + P \in \mathcal{O}(n^2)$ parameters to estimate. As a result, the number of parameters grows as a linear function of the lag order.

Compact form

Vector and matrix operations can be readily generalized to the tensor case, but notions of tensor algebra are necessary to proceed. See Appendix A.2 for some basics of tensor notation and calculus. For a comprehensive explanation see (Cichocki, 2018; Kolda and Bader, 2009).

Although an estimation procedure for the MAR(1) is readily available (Chen et al., 2021), a compact form of the model is instrumental for developing a more general and coherent estimation procedure of the MAR(P). We therefore derive a comprehensive compact form for any general K-dimensional TAR(P), which admits the MAR(P) and MAR$^*(P)$ as special cases, by establishing a connection with the general Tensor Linear Regression (TLR) model. Indeed, as in its compact form the VAR emerges as a multivariate linear regression (MLR) model, the analogue of the MAR is a TLR model.

The compact form of the VAR can be easily derived by creating the matrix of dependent variables $\mathbf{Y}$ by horizontally stacking each $\mathbf{y}_t$ for $t = 1, ..., T$. However, for each time point, the variables in a MAR are already matrix shaped, so that its related compact form turns out to be the a third order tensor, the closest generalization of a matrix.

Let us consider a K way (order) tensor $\mathcal{Y} \in \mathbb{R}^{J_1,...,J_K}$ which is a K-dimensional array with entries $\mathcal{Y}_{j_1,...,j_K}$ with $j_k = 1, ..., J_k$ for $k = 1, ..., K$. We define the response and explanatory tensor for the MAR in equation (3.1). Let $\mathcal{Y}$ be a $K = 3$ way response tensor of dimension $[G \times N \times T - P]$ with $\mathcal{Y}_{:,:,t} = \mathbf{Y}_t$. Then, define the explanatory tensor $\mathcal{X}$ to be of dimension

$[GP \times NP \times T-P]$. When fixing the third dimension $j_3 = t = 1, ..., T-P$, we obtain tensor slices of dimensions $[GP \times NP]$, which we fill with the lagged values of $\mathbf{Y}_t$ and zeros otherwise:

$$\mathcal{X}_{:,:,t} = \begin{bmatrix} \mathbf{Y}_{t-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{t-2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Y}_{t-P} \end{bmatrix}.$$

In the case of the the MAR$^*(P)$ in equation (3.3), the response and explanatory objects of interest need to be slightly modified. We define $\mathcal{X}$ to be a $K = 4$ order tensor of dimensions $[G \times N \times P \times T - P]$ such that $\mathcal{X}_{:,:,i,t} = \mathbf{Y}_{t-i}$ and, for the sake of coherency, we modify $\mathcal{Y}$ to be a four-dimensional tensor $\mathcal{Y} \in \mathbb{R}^{G \times N \times 1 \times T-P}$ as well.

We may now write a unique compact form which encompasses both MAR(P) and MAR$^*(P)$. Let $\mathcal{B} = \{[\mathbf{A}_1, ..., \mathbf{A}_P], [\mathbf{B}_1, ..., \mathbf{B}_P], \mathbf{I}_{T-P}\}$ and $\mathbf{\Sigma} = \{\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{I}_{T-P}\}$ in case of MAR(P), $\mathcal{B} = \{\mathbf{A}, \mathbf{B}, \mathbf{c}, \mathbf{I}_{T-P}\}$ and $\mathbf{\Sigma} = \{\mathbf{\Sigma}_1, \mathbf{\Sigma}_2, \mathbf{\Sigma}_3, \mathbf{I}_{T-P}\}$[2] in case of MAR$^*(P)$. If $\bar{\times}$ is the Tucker product and $\mathcal{U}$ is a tensor white noise, then we have

$$\begin{aligned} \mathcal{Y} &= \mathcal{X} \bar{\times} \mathcal{B} + \mathcal{E} \\ \mathcal{E} &= \mathcal{U} \bar{\times} \mathbf{\Sigma}^{1/2} \\ \mathcal{U} &\sim \mathcal{MN}(\mathbf{0}, \mathbf{I}_{J_1}, ..., \mathbf{I}_{J_K}) \end{aligned} \tag{3.4}$$

Notice that equation (3.4) can be seen as the multilinear generalization of the compact form of VAR models–see Lütkepohl (2005).

It will be convenient to derive the MLR representation of the models. To simplify the notation a little, let $\mathbf{\Phi}_1 = [\mathbf{A}_1, ..., \mathbf{A}_P]$ and $\mathbf{\Phi}_2 = [\mathbf{B}_1, ..., \mathbf{B}_P]$ for the MAR(P), $\mathbf{\Phi}_1 = \mathbf{A}$, $\mathbf{\Phi}_2 = \mathbf{B}$ and $\mathbf{\Phi}_3 = \mathbf{c}$ for the MAR$^*(P)$. Then, let $\tilde{\mathbf{Y}}^k = \text{mat}_k(\mathcal{Y})\mathbf{\Sigma}_{-k}$, $\tilde{\mathbf{X}}^k = \text{mat}_k(\mathcal{X})\mathbf{\Sigma}_{-k}$ and $\tilde{\mathbf{E}}^k = \text{mat}_k(\mathcal{E})\mathbf{\Sigma}_{-k}$, where

[2]Notice that, being the third dimension of $\mathcal{Y}$ of order 1, $\mathbf{\Sigma}_3 = \sigma_3^2$ collapses to a scalar.

$\boldsymbol{\Sigma}_{-k} = \boldsymbol{\Sigma}_K \otimes ... \otimes \boldsymbol{\Sigma}_{k+1} \otimes \boldsymbol{\Sigma}_{k-1}, ..., \boldsymbol{\Sigma}_1$, where $\text{mat}_k(\cdot)$ denotes the k-mode matricization operator as defined in Appendix A.2. By matricizing both the sides of equation (3.4) for each k, one can easily derive a MLR model which highlights the kth way conditional mean and variance matrices $\{\boldsymbol{\Phi}_k, \boldsymbol{\Sigma}_k\}$:

$$\begin{aligned} \tilde{\mathbf{Y}}^k &= \boldsymbol{\Phi}_k \tilde{\mathbf{X}}^k + \tilde{\mathbf{E}}^k \\ \tilde{\mathbf{E}}^k &\sim \mathcal{MN}(\mathbf{0}, \boldsymbol{\Sigma}_k, \mathbf{I}_{J_{-k}}) \end{aligned} \tag{3.5}$$

The latter formulation allows the estimation to be carried out for each of the K modes separately. Notice that in the case of vector-valued time series, the compact form in equation (3.4) reduces to the formulation in equation (3.5).

Shrinkage Priors

Consider the compact form in (3.4). Let $J_{-k} = \prod_{l \neq k} J_l$ so that the likelihood can be written as

$$\mathcal{L}(\mathcal{Y}|\mathcal{B}, \boldsymbol{\Sigma}) \propto \prod_{k=1}^{K} |\boldsymbol{\Sigma}_k|^{-\frac{J_{-k}}{2}} \exp\left(-\frac{1}{2}||(\mathcal{Y} - \mathcal{X} \bar{\times} \mathcal{B}) \bar{\times} \boldsymbol{\Sigma}^{-\frac{1}{2}}||_2^2\right). \tag{3.6}$$

Also, notice that the likelihood for the MAR model in terms of $\{\boldsymbol{\Phi}_k, \boldsymbol{\Sigma}_k\}$, given the other parameters, is proportional to the likelihood in (3.6):

$$\mathcal{L}(\mathcal{Y}|\boldsymbol{\Phi}_k, \boldsymbol{\Sigma}_k) \propto \mathcal{L}(\mathcal{Y}|\boldsymbol{\Phi}, \boldsymbol{\Sigma}). \tag{3.7}$$

This implies that not only ML but also Bayesian estimation can be carried for each dimension separately (see, e.g., Hoff, 2015). As a consequence, sparsity can be induced in the two modes independently, which eases the introduction of regularization methods for MAR models.

We will consider a spike-and-slab framework for the prior to iteratively induce sparsity in the two mode-specific coefficients of the MAR. A related approach has been developed for VAR models (George et al., 2008). For

each coefficient $\phi_{i,k} \in \boldsymbol{\phi}_k = \text{vec}(\boldsymbol{\Phi}_k)$, we have a binary indicator $\gamma_{i,k} \in \{0, 1\}$, which encodes the state of $\phi_{i,k}$ (the "spike" inactive state for $\gamma_{i,k} = 0$ and the "slab" active state for $\gamma_{i,k} = 1$). Given $\gamma_{i,k}$, the conditional mixture prior for each $\phi_{i,k}$ can be expressed as

$$\phi_{i,k}|\gamma_{i,k} \sim (1 - \gamma_{i,k})\mathcal{N}(0, \tau_0) + \gamma_{i,k}\mathcal{N}(0, \tau_1), \tag{3.8}$$

which is controlled by the two hyperparameters τ_0 and τ_1 By selecting the two such that former approaches 0 whereas the latter is arbitrarly large, $\gamma_{i,k}$ is able to identify restrictions on $\phi_{i,k}$. The prior for the kth mode conditional mean parameters can be rewritten compactly as:

$$\boldsymbol{\phi}_k|\boldsymbol{\gamma}_k \sim \mathcal{N}(\mathbf{0}, \mathcal{V}_k) \tag{3.9}$$

where $\mathcal{V}_k = \text{diag}(v_{1,k}, ..., v_{n_k,k})$ and $v_{i,k} = (1 - \gamma_{i,k})\tau_0 + \gamma_{i,k}\tau_1$, being n_k the cardinality of $\boldsymbol{\phi}_k$. We assume each $\gamma_{i,k}$ is independent Bernoulli, i.e. :

$$\gamma_{i,k}|\theta_k \sim \mathcal{B}\text{er}(\theta_k). \tag{3.10}$$

With a priori information on the level of sparsity in the coefficients, one can set θ_k accordingly. Therefore, we endow each indicator with a Beta-Bernoulli hierarchical prior scheme:

$$\begin{aligned} \pi(\boldsymbol{\gamma}_k|\theta_k) &= \theta_k^{|\gamma_k|}(1 - \theta_k)^{n_k - |\gamma_k|} \\ \theta_k &\sim \mathcal{B}\text{eta}(\alpha_k, \beta_k) \end{aligned} \tag{3.11}$$

where $|\gamma_k| = \sum_i \gamma_{i,k}$.

Notice that the two covariance matrices of the MAR enter the likelihood in a multiplicative way, meaning their scales are not separately identifiable. Without imposing restrictions, such quantities would be determined completely by the prior covariance matrices. Further restrictions on the scales are therefore required without any additional a priori information.

As in Hoff (2011, 2015), we introduce dependence between the Inverse Wishart prior distribution of each $\boldsymbol{\Sigma}_k$ by adding a level of hierarchy through a hyperparameter ξ

$$\begin{aligned} \xi &\sim \mathcal{G}a(\eta_1, \eta_2), \\ \boldsymbol{\Sigma}_k|\xi &\sim \mathcal{W}^{-1}(\xi\Omega_k, \nu_k), \end{aligned} \tag{3.12}$$

such that by setting $\Omega_k = \mathbf{I}_{J_k}/J_k$ and $\nu_k = J_k + 2$ the total variance is controlled only by a K-th power of ξ:

$$\mathrm{E}\Big[\prod_k \mathrm{tr}(\boldsymbol{\Sigma}_k)\Big] = \xi^K. \tag{3.13}$$

Thus, if we let $\boldsymbol{\Delta}_k$ be the collection of all the k-th mode parameters except for $\boldsymbol{\gamma}_k$, the joint prior distribution takes the form

$$\pi(\boldsymbol{\Delta}_1, ..., \boldsymbol{\Delta}_K, \boldsymbol{\gamma}_1, ..., \boldsymbol{\gamma}_K) = \prod_k \pi(\boldsymbol{\phi}_k|\boldsymbol{\gamma}_k)\pi(\boldsymbol{\gamma}_k|\theta_k)\pi(\theta_k)\pi(\boldsymbol{\Sigma}_k|\xi)\pi(\xi). \tag{3.14}$$

3.3 Bayesian Estimation

We develop two computational methods for fitting the proposed Bayesian model (i) a Gibbs sampler and (ii) a maximum a posteriori (MAP) estimation procedure via EMVS. The Gibbs sampling algorithm will produce more accurate estimates and allow estimation of the full posterior, however, it is expected to be slower than the EMVS procedure, which aims only at identification of posterior modes by iteratively maximizing the conditional expectation of the log posterior.

An outline of the proposed Gibbs sampling procedure is given in Algorithm 10. We report details on the full conditional posterior distributions in Appendix A.2. Note that while we defer description of how we obtain the MAP estimates, this is what we propose to use as starting values for the Gibbs sampler. Posterior modes are often good starting values to use in MCMC simulation experiments (Geyer, 2010; Jones and Qin, 2022; Vats et al., 2021).

Algorithm 1: MCMC

Starting values: MAP estimate $\hat{\mathbf{\Phi}}_k$, $\hat{\mathbf{\Sigma}}_k$.
Hyperparameters: $\tau_0, \tau_1, \alpha_k, \beta_k, \Omega_k, \nu_k$.
Initialize : $\mathbf{\Phi}_k^{[0]} = \hat{\mathbf{\Phi}}_k$, $\mathbf{\Sigma}_k^{[0]} = \hat{\mathbf{\Sigma}}_k$, $\theta_k^{[it]} = \alpha_k/\beta_k$,
n. of iterations MC, size of Burn in BU.

for $j = 1$ **to** $MC + BU$ **do**
  for $k = 1$ **to** K **do**
    for $i = 1$ **to** n_k **do**
      Draw $\gamma_{i,k}^{[j]}$ from the Bernoulli distribution (A.11).
    Compute $\tilde{\mathbf{Y}}_k$ and $\tilde{\mathbf{X}}_k$ with $\mathbf{\Phi}_k^{[j-1]}$ and $\mathbf{\Sigma}^{[j-1]}$ as in subsection A.2.
    Draw $\boldsymbol{\phi}_k^{[j]}$ from the multivariate Normal distribution (A.13).
    Draw $\mathbf{\Sigma}_k^{[j]}$ from the Inverse Wishart distribution (A.15).
    Draw $\theta_k^{[j]}$ from the Beta distribution (A.16).
  Draw $\xi^{[j]}$ from the Gamma distribution (A.17).
  Compute $\mathcal{B}^{[j]}$ and $\mathbf{\Sigma}^{[j]}$ with $\mathbf{\Phi}_1^{[j]}, ..., \mathbf{\Phi}_K^{[j]}$ and $\mathbf{\Sigma}_1^{[j]}, ..., \mathbf{\Sigma}_k^{[j]}$ and renormalize via (3.18)).

MAP Estimation

A global optimization procedure to find the posterior mode can be separately set for each order K. However, given the mixture of prior for each $\boldsymbol{\phi}_k$, direct optimization of the log conditional posterior $\log \pi(\mathbf{\Delta}_k|\mathcal{Y})$ has no analytical solution. The presence of the sum prevents the logarithm from acting directly on the joint conditional posterior, which results in complicated expressions for the MAP solution.

We employ an Expectation Conditional Maximization (ECM) algorithm (Meng and Rubin, 1993), which indirectly maximizes $\log \pi(\mathbf{\Delta}_k, \boldsymbol{\gamma}_k|\mathcal{Y})$ by iteratively maximizing its expected value under the posterior distribution of the latent variable. At iteration j, this expectation, denoted by $\mathcal{Q}(\mathbf{\Delta}_k^{[j]}|\mathbf{\Delta}_k^{[j-1]})$, is given by

$$\mathcal{Q}(\mathbf{\Delta}_k|\mathbf{\Delta}_k^{[j-1]}) = \mathrm{E}_{\boldsymbol{\gamma}_k|\mathbf{\Delta}_k^{[j-1]}}\Big[\log \pi(\mathbf{\Delta}_k, \boldsymbol{\gamma}_k|\mathcal{Y})|\mathbf{\Delta}_k^{[j-1]}\Big], \tag{3.15}$$

which constitutes the E-step of the algorithm. In the M-step, we derive the

revised estimate of all the other parameters by maximizing the function:

$$\boldsymbol{\Delta}_k^{[j]} = \underset{\boldsymbol{\Delta}_k}{\operatorname{argmax}} \ \mathcal{Q}(\boldsymbol{\Delta}_k|\boldsymbol{\Delta}_k^{[j-1]}) \tag{3.16}$$

where $\mathcal{Q}(\boldsymbol{\Delta}_k|\boldsymbol{\Delta}_k^{[j-1]})$ can be decomposed as:

$$\begin{aligned}
\mathcal{Q}_{1,k}(\boldsymbol{\phi}_k^{[j]}|-) =& -\frac{1}{2}\Big[(\boldsymbol{\phi}_k - \hat{\boldsymbol{\phi}}_k)'(\tilde{\mathbf{X}}^{k'}\tilde{\mathbf{X}}^k \otimes \boldsymbol{\Sigma}_k^{-1})(\boldsymbol{\phi}_k - \hat{\boldsymbol{\phi}}_k) + \boldsymbol{\phi}_k^{'}\mathcal{V}_k^{-1}\boldsymbol{\phi}_k^{'}\Big] \\
\mathcal{Q}_{2,k}(\boldsymbol{\Sigma}_k^{[j]}|-) =& -(J_k + J_{-k} + \nu_k + 1)\log|\boldsymbol{\Sigma}_k| \\
& - \operatorname{tr}\Big\{\boldsymbol{\Sigma}_k^{-1}\Big[\xi\Omega_k + (\tilde{\mathbf{Y}}^k - \boldsymbol{\Phi}_k\tilde{\mathbf{X}}^k)(\tilde{\mathbf{Y}}^k - \boldsymbol{\Phi}_k\tilde{\mathbf{X}}^k)'\Big]\Big\} \\
\mathcal{Q}_{3,k}(\theta_k^{[j]}|-) =& (|\boldsymbol{\gamma}_k| + \alpha_k - 1)\log(\theta_k) + (n_k - |\boldsymbol{\gamma}_k| + \beta_k - 1)\log(1-\theta_k) \\
\mathcal{Q}_4(\xi^{[j]}|-) =& \frac{1}{2}\sum_k J_k\nu_k\log(\xi) - \frac{1}{2}\sum_k \operatorname{tr}(\Omega_k\boldsymbol{\Sigma}_k^{-1})\xi + (\eta_1 - 1)\log(\xi) - \eta_2\xi
\end{aligned} \tag{3.17}$$

Notice that there is an identifiability issue arising from the structure of the MAR. Given the properties of the Kronecker product, if $\boldsymbol{\Phi}_1, ..., \boldsymbol{\Phi}_K$ are a solution of the problem, so are $f_1\boldsymbol{\Phi}_1, ..., f_K\boldsymbol{\Phi}_K$, with the condition $\prod_k f_k = 1$. To keep iterations of both the Gibbs and the EMVS stable, we choose f_k such that the magnitude between the various parameter matrices remains as stable as possible in the following way:

$$f_k = \frac{\prod_{l\neq k}||\boldsymbol{\Phi}_l||^{\frac{1}{K}}}{||\boldsymbol{\Phi}_k||^{\frac{K-1}{K}}} \tag{3.18}$$

The same applies for the covariance matrices. We illustrate the complete EMVS estimation procedure in Algorithm 13.

3.4 Simulations

We design two simulation experiments. The first one is aimed at studying the small sample efficiency of our proposed method. The second one evaluates the estimation error, forecasting performance, and computational time of

Algorithm 2: EMVS

Starting values: ML estimate $\hat{\mathbf{\Phi}}_k$, $\hat{\mathbf{\Sigma}}_k$.
Hyperparameters: $\tau_0, \tau_1, \alpha_k, \beta_k, \Omega_k, \nu_k$.
Initialize : $\mathbf{\Phi}_k^{[0]} = \hat{\mathbf{\Phi}}_k$, $\mathbf{\Sigma}_k^{[0]} = \hat{\mathbf{\Sigma}}_k$, $\theta_k^{[it]} = \alpha_k/\beta_k$, $j = 0$, tolerance ϵ.

while $Tol > \epsilon$ **do**
  $j = j + 1$.
  for $k = 1$ **to** K **do**
    for $i = 1$ **to** n_k **do**
      Compute $\mathrm{E}_{\gamma_k|.}(\gamma_{i,k})$ from eq. (A.18).
      Compute $\mathrm{E}_{\gamma_k|.}(v_{i,k}^{-1})$ from eq. (A.19).
    Compute $\tilde{\mathbf{Y}}_k$ and $\tilde{\mathbf{X}}_k$ as in subsection 3.2.
    Update $\phi_k^{[j]}$ from eq. (A.20).
    Update $\mathbf{\Sigma}_k^{[j]}$ from eq. (A.21).
    Update $\theta_k^{[j]}$ from eq. (A.22).
  Update $\xi^{[j]}$ from eq. (A.23).
  Compute $\mathcal{B}^{[j]}$ and $\mathbf{\Sigma}^{[j]}$ with $\mathbf{\Phi}_1^{[j]}, ..., \mathbf{\Phi}_K^{[j]}$ and $\mathbf{\Sigma}_1^{[j]}, ..., \mathbf{\Sigma}_k^{[j]}$ and renormalize via (3.18).
  Compute $Tol = \max||\mathbf{\Phi}_k^{[j]} - \mathbf{\Phi}_k^{[j-1]}||_2^2$.

our proposed approach, relative to multiple existing estimation methods for longitudinal data. We further perform a comparative convergence analysis of the MCMC in Appendix A.2.

We set our hyperparameters to be $\tau_0 = 0.01$, $\tau_1 = 4$, $\Omega_k = \mathbf{I}_{J_k}/J_k$, $\nu_k = J_k + 2$, $\alpha_k = 1$, $\beta_k = J_k - 1$, $\nu_1 = 1$, and $\nu_2 = 1$ for both experiments. In the second experiment, we run all the Gibbs samplers for 1000 iterations. Given a confidence level of 0.05, this number of simulations ensures stable estimation results, being above the minimum effective sample size (mESS) at a tolerance of 0.05 in the largest scenario (Vats et al., 2019). Details on the convergence of the sampler of our proposed model can be found in Appendix A.2.

Small Sample Efficiency

We compare the small sample efficiencies of ML and MAP estimators by letting the length of the time series T and the level of sparsity in the autoregressive coefficients vary. The main purpose of this experiment is to obtain qualitative understanding of the small sample covariances of the two estimators under different sparsity settings. We use the MAP estimator for comparison purposes given its computational convenience.

To this aim, we simulate our synthetic data as follows. We generate matrix-valued time series observations from a MAR(1) with dimensions $G, N = 8$, with different lengths of the time series $T = 50, ..., 1000$. We place true nonzero loadings in the left and right model coefficient matrices according to four different settings. In settings $i = 1, ..., 4$ we will place $SP_i = 1, 2, 4, 8$ non-zero coefficients for each equation both in $\mathbf{A}$ and $\mathbf{B}$, respectively. Covariance matrices are set to $\mathbf{\Sigma}_1 = \mathbf{I}_G$ and $\mathbf{\Sigma}_2 = \mathbf{I}_N$. Our data generating process (DGP) is such that the main diagonal blocks of $\mathbf{A}$ and $\mathbf{B}$ are $[SP_i \times SP_i]$ matrices whose elements are drawn from a $\mathcal{N}(0, 1)$, and zero otherwise. Notice that with $i = 1$ and $i = 4$ the two coefficient matrices are diagonal and full, respectively. In setting our MAP estimator priors, we fix $\alpha_k = 1$, and choose $\beta_k = J_k - 1$ so to reflect a prior belief of a sparse DGP as the one in setting 1. Estimation errors are measured by the mean squared error (MSE):

$$MSE(\hat{\mathbf{\Phi}}) = \frac{\text{tr}\Big[(\mathbf{\Phi} - \hat{\mathbf{\Phi}})'(\mathbf{\Phi} - \hat{\mathbf{\Phi}})\Big]}{G \times N} \tag{3.19}$$

where for the MAR model $\mathbf{\Phi} = \mathbf{B} \otimes \mathbf{A}$ and $\hat{\mathbf{\Phi}} = \hat{\mathbf{B}} \otimes \hat{\mathbf{A}}$.

Figure 6 illustrates a comparison of the small sample efficiencies of the ML and MAP estimators, as measured by the average estimation errors over 50 repetitions of $MSE(\hat{\mathbf{\Phi}})$. The figure shows a decreasing trend of the average error of both estimators as T grows. As expected, given the imposed prior beliefs the more sparse the DGP, the more the MAP estimator results in a higher efficiency in small sample. This is magnified in setting 1 (DGP with

diagonal coefficient matrices), while the difference in efficiency between the two estimators gradually vanishes as the number of non-zero coefficients grow. Notice however that, even in setting 4 (DGP with full coefficient matrices) the efficiency of the two estimators is still comparable.

Comparative Estimation, Forecasting and Computational Performances

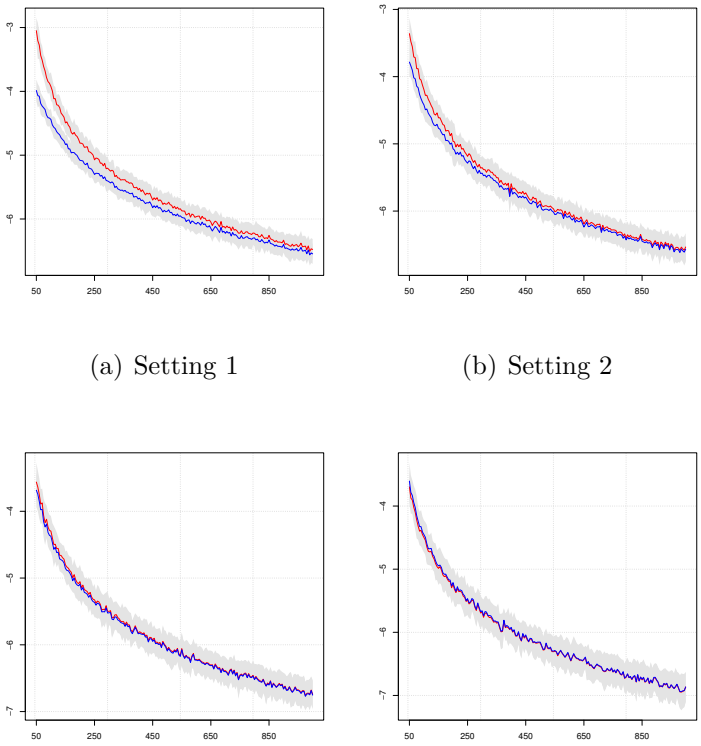
We aim to compare the estimation error, forecasting performance, and computation time of the illustrated estimators with other relevant competing alternatives for panel data. This is done for different choices of the matrix dimensions $(G, N) = (2, 3), (4, 6), (8, 10)$, so setup a "small", a "medium" and a "large" setting, relative to a reference sample size $T = 100$. We consider the three MAR estimators: ML, MCMC (Bayes), and MAP. The alternatives considered are the stacked VAR estimator (VAR), the country-block panel VAR (CB), the Cross-sectional Shrinkage (CC) approach of Canova and Ciccarelli (2009, 2013), the Stochastic Search Specification Selection (SSSS) of Koop and Korobilis (2016). We briefly describe the competing alternative models along with their hyperparameter specification in Appendix A.2.

We simulate a sparse VAR(1) able to reflect recurrent patterns in multi-country and multi-variable applications. Let $\mathbf{\Xi}$ be the $[GN \times GN]$ matrix of autoregressive coefficients and $\mathbf{\Xi}_{i,j} \in \Xi$ be the $[G \times G]$ country j block of parameters in the country i equations, and consider its entries related to the indicators k, l:

$$\mathbf{\Xi}_{i,j}^{k,l} = \begin{cases} \lambda \lambda_n^{|i-j|} \lambda_g^{|k-l|} & \text{if } |i-j| \leq r, |k-l| \leq r \\ 0 & \text{otherwise} \end{cases} \tag{3.20}$$

where $\lambda = 0.95$ is an overall constant term, λ_n and λ_g are a country and an indicator penalty term, respectively, such that $\lambda_n = 0.5 + \mathcal{U}_{[-0.1,0.1]}$ and $\lambda_g = 0.5 + \mathcal{U}_{[-0.1,0.1]}$, and $r = 2$. The DGP is able to reflect heterogeneities in both dimensions such that coefficients are affected by a country penalty

Figure 6: Efficiency comparison between ML and MAP



(a) Setting 1

(b) Setting 2

(c) Setting 3

(d) Setting 4

Comparison of the efficiencies of ML (red) and MAP (blue) estimators over 50 repetitions for $T = 50, ..., 1000$ under four sparsity settings. The figure shows the logarithm of the $MSFE(\hat{\mathbf{\Phi}})$ (y axis) over different T (x axis). Gray shaded areas represent the 1 standard deviation confidence bounds.

λ_n, an indicator penalty λ_g and both of them combined. Such penalties act when the row and column distances $|i-j|$ and $|k-l|$ of the elements of the matrix $\Xi_{i,j}^{k,l}$ do not exceed the threshold level r.

Estimation errors of the models are measured by the MSE as in equation (3.19), whereas the forecasting performance for fixed forecast horizon H is assessed by means of the Mean Squared Forecast Error (MSFE):

$$MSFE(H) = \frac{1}{G \times N \times H} \sum_{h=1}^{H} (\hat{\mathbf{y}}_{T+h|T} - \mathbf{y}_{T+h})' (\hat{\mathbf{y}}_{T+h|T} - \mathbf{y}_{T+h}), \quad (3.21)$$

where $\hat{\mathbf{y}}_{T+h|T}$ is the H-step forecast obtained with information up to the last sample size T.

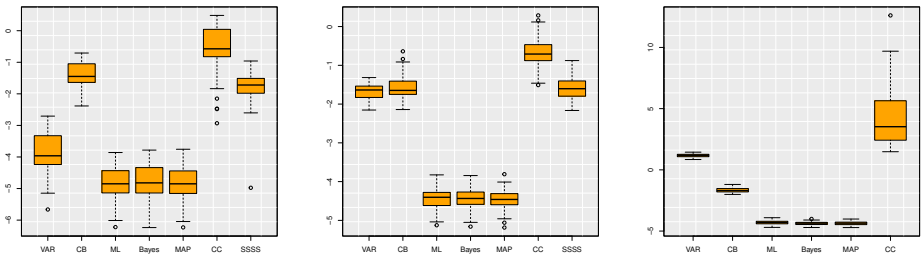
We illustrate in Figure 7 numerical results on the model MSE and MSFE related to $H = 1$. When the estimation of some competitor model was infeasible, corresponding results are omitted. We report results related to repeated experiments with $T = 500$ and a sparsity setting generated by $r = 3$ in Appendix A.2.

One common finding is that the proposed estimators, along with ML, tend to overperform all competing alternatives, both in terms of estimation and forecasting performances. The difference in performances becomes more substantial when considering the "large" dimensional setting. Results with $T = 500$ and $r = 3$ are consistent with the expectations that as T grows and true coefficient matrices are less sparse, performance gains compared to a standard VAR diminish in the "small" dimensional setting.

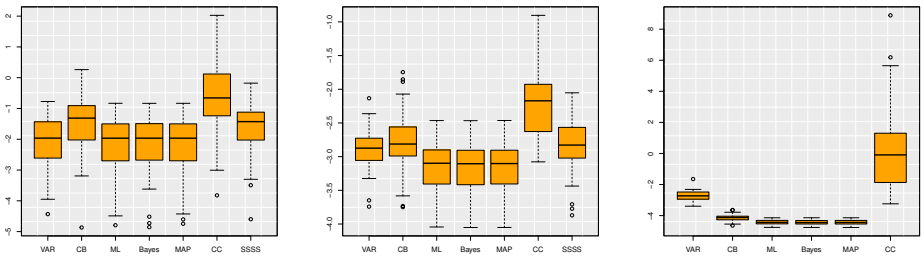
We now report the average computational time over 50 repetitions for the three different combination of G and N and $T = 100$ in Table 3. Being the estimators of VAR and CB closed form, their related execution times are naturally the lowest.

As expected, the computational times of MAP are much more favorable than those of the MCMC. The difference between the two lies in the fact that while the MCMC requires thousands of iterations to approximate the

Figure 7: Simulation performances



(a) G=2, N=3 (b) G=4, N=6 (c) G=8, N=10



(d) G=2, N=3 (e) G=4, N=6 (f) G=8, N=10

Average estimation error (MSE) (a,b,c) and 1 step ahead forecasting performance (MSFE) (d,e,f) over 50 repetitions of each model for $(G, N) = (2, 3), (4, 6), (8, 10)$ with $T = 100$.

Table 3: Computational time

	G=2, N=3	G=4, N=6	G=8, N=10
VAR	8.7×10^{-5} (1.2×10^{-4})	1.1×10^{-4} (3.8×10^{-4})	4.7×10^{-4} (2.1×10^{-4})
CB	8.2×10^{-5} (1.5×10^{-4})	1.6×10^{-4} (7.3×10^{-4})	3.1×10^{-4} (2.1×10^{-4})
ML	4.6×10^{-4} (8.7×10^{-3})	0.01 (4.5×10^{-3})	0.01 (0.01)
Bayes	9.98 (0.51)	30.9 (10.1)	85.72 (23.24)
MAP	0.01 (7.2×10^{-3})	0.03 (0.02)	0.07 (0.02)
CC	0.26 (0.02)	1.82 (0.55)	25.67 (3.78)
SSSS	5.18 (0.24)	368.52 (76.58)	-

Average computational times over 50 repetitions for $(G, N) = (2, 3), (4, 6)$ and $(8, 10)$, with $T = 100$. Standard deviations are shown in parentheses.

joint posterior distribution, the EMVS algorithm reaches convergence with a lower number of iterations, usually less than 100. Despite this difference, the two have exhibited comparable performances in estimating the autoregressive coefficient and forecasting performances.

In our settings, the MCMC is also generally slower than CC, which exploits factors to reduce dimensionality, but is much faster than SSSS in "medium" and "large" settings. Further, given the dimensionality reduction achieved by the MAR, computation times of the MCMC grow slower with increasing model parameters than those related to the CC and SSSS estimators.

Notice that the computational times of MAP are distinctly lower, not only than those of its full Bayesian counterpart, but also than those related to the CC and SSSS estimators. Notwithstanding this, estimation and forecasting performances of MAP (and, in general, of the two proposed computational methods) are generally superior in high-dimensional settings.

MCMC convergence analysis

Here we assess and compare the convergence of the competing models' MCMC algorithms by performing a diagnostic based on the criteria of Gong and Flegal (2016) and Vats et al. (2019). In other words, we want to determine when should sampling stop in order to get adequate parameter estimates. Given a level of confidence α and tolerance ϵ, one simple strategy is to run the Markov Chain at least for a number of iterations larger than the minimum Effective Sample Size (mESS). In this context, Vats et al. (2019) propose a multivariate framework for terminating simulation in MCMC and define the multivariate mESS as:

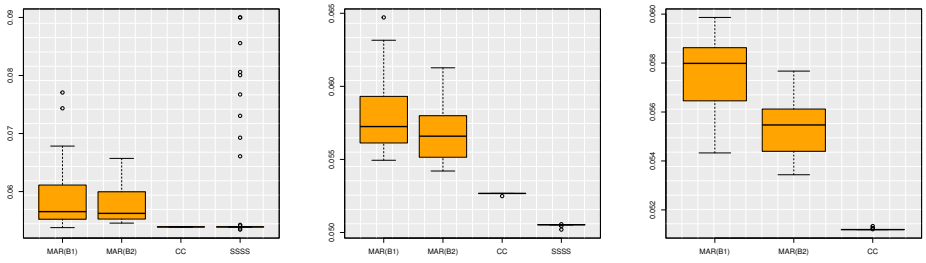
$$mESS \geq \frac{2^{2/p}\pi}{[p\Gamma(p/2)]^{2/p}} \frac{\chi^2_{1-\alpha,p}}{\epsilon^2} \tag{3.22}$$

where p is the number of parameters to be estimated. It should be noticed that $mESS$ is a function of p, α and ϵ and is therefore independent of the Markov chain or the underneath process. This paves the way to model comparison in terms of MCMC convergence analysis.

Notice that the three Gibbs samplers employed in this study are characterized by a different number of parameters p. In particular, for each combination of $[G, N]$ of the simulation study, the Gibbs sampler for the MAR produces two chains of dimensions $[G^2 \times MC]$ and $[N^2 \times MC]$, respectively. The samplers for the CC and the SSSS procedures produce two chains of dimensions $[GNf \times MC]$ (with $f < GN$) and $[G^2N^2 \times MC]$. Moreover, the CC is a factor model, for which the number of factors can differ among datasets, and is sensitive to the statistical method employed to extract the relevant factors. As a consequence, it is not possible to choose a common number of simulations such that the tolerance ϵ is kept constant among all settings and models.

However, a possible strategy to overcome this issue is to use equation (3.22) in another way. Instead of choosing a different value of MC for each model and setting, one can set an arbitrarily large level so that it is guaran-

Figure 8: Tolerance



(a) G=2, N=3 (b) G=4, N=6 (c) G=8, N=10

Tolerance ϵ over 50 repetitions of each model for $(G, N) = (2, 3), (4, 6), (8, 10)$ with $T = 100$.

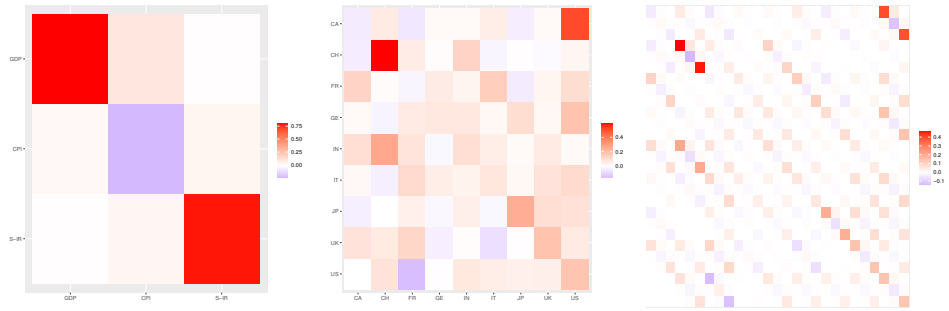
teed that in all settings at least a minimum pre-determined ϵ is reached, and then compare the samplers in terms of tolerance levels. We can determine a model specific estimated ESS, calculated after running the MCMC, and then get the corresponding tolerance level ϵ via equation (3.22). Given MC iterations in a Markov chain, the ESS measures the size of an i.i.d. sample with the same standard error. In a multivariate setting, the ESS is given by:

$$ESS = MC\left(\frac{|\mathbf{\Lambda}|}{|\mathbf{\Sigma}|}\right)^{1/p} \tag{3.23}$$

where $\mathbf{\Lambda}$ is the sample covariance matrix and $\mathbf{\Sigma}$ is an estimate of the variance of the asymptotic normal distribution. Replacing $mESS$ in eq. (3.22) with ESS, we can express it in terms of the tolerance ϵ, which can be viewed as a comparison in terms of convergence of the Gibbs sampler. The smaller the minimum effective samples, the larger the tolerance, and hence the smaller the number of simulations required.

A visual comparison of the tolerance level ϵ for the three models under the different setting is depicted in Figure 8. The figure shows that our

Figure 9: Posterior median of autoregressive coefficients



(a) $\hat{\mathbf{A}}$ (b) $\hat{\mathbf{B}}$ (c) $\hat{\mathbf{B}} \otimes \hat{\mathbf{A}}$

Median of the posterior entries of the first order left coefficient matrix $\hat{\mathbf{A}}$ (a), of the right one $\hat{\mathbf{B}}$ (b), and of $\hat{\mathbf{B}} \otimes \hat{\mathbf{A}}$ (c).

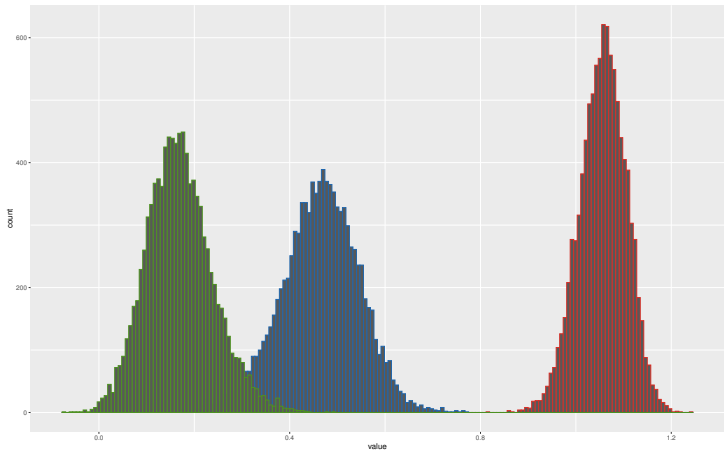
sampler achieves generally greater tolerance than the analyzed competing alternatives, particularly in high-dimensional settings.

3.5 Application: Panel of Country Economic Indicators

We now conduct an empirical application of the proposed model to a panel of $G = 9$ world countries, which currently represent approximately the 64% of the world total Gross Domestic Product (GDP): Canada (CA), China (CH), France (FR), Germany (GE), India (IN), Italy (IT), Japan (JP), United Kingdom (UK) and United States (US). We consider quarterly observations of $N = 3$ economic indicators: GDP, Consumer Price Index (CPI) and Short-term Interest Rates (S-IR), all expressed in log differences. The sample ranges from 1980Q1 to 2019Q4 ($T = 162$). The data is inherently multidimensional, as observations are generated in matrix form, where rows represent indicators and columns countries.

We estimate a fully Bayesian MAR*(3), which yields the full conditional

Figure 10: Posterior of the temporal coefficient vector



Posterior distribution of $\hat{c}_1$ (red line), $\hat{c}_2$ (blue line), $\hat{c}_3$ (green line).

posterior distribution of the parameters of interest. We illustrate in Figure 9 the posterior median of the left ($\hat{\mathbf{A}}$) and right ($\hat{\mathbf{B}}$) coefficient matrices, along with the reconstructed coefficient matrix $\hat{\mathbf{B}} \otimes \hat{\mathbf{A}}$. Note that, given that only the Kronecker product $\hat{\mathbf{B}} \otimes \hat{\mathbf{A}}$ is uniquely identified, only magnitudes related to the left and right coefficient matrices can be meaningfully interpreted, rather than signs.

The figure shows that, as intuition suggests, the diagonal elements of the parameter matrices concur to a large portion of the system's autoregressive dynamics. This is more evident from the left coefficient matrix, indicating a strong autocorrelation in the variable dimension. From the first order right coefficient matrix estimate notice that the two largest impacts are those of China on itself and US on Canada. While the former can be explained through the relatively low impact of other countries on the Chinese economy as a whole, the latter seems reasonable given the geographical proximity and large trade activities between them.

Figure 10 shows the posterior distribution of the lag order coefficients $\hat{c}$.

Table 4: Average MSFE and computational time

	Av. Log MSFE	Av. Computational Time
VAR	4.41×10^{-5} (2.77×10^{-5})	1.91×10^{-4} (9.97×10^{-5})
CB	3.84×10^{-5} (1.96×10^{-5})	1.72×10^{-4} (7.44×10^{-5})
ML	3.21×10^{-5} (1.85×10^{-5})	1.83×10^{-2} (7.00×10^{-3})
Bayes	3.09×10^{-5} (1.79×10^{-5})	25.88 (5.10)
MAP	3.11×10^{-5} (1.79×10^{-5})	0.17 (0.07)
CC	4.40×10^{-5} (2.85×10^{-5})	1.9 (0.59)
SSSS	3.83×10^{-5} (1.75×10^{-5})	776.06 (51.82)

Average Logarithm of the MSFE and average computational time over 50 repeated estimations of three different MAR estimators (ML, Bayes, MAP) against competing alternatives. Standard deviations are shown in parentheses.

It is clear how values of the posterior distribution of $\hat{c}$ decrease with the lag order itself, as one could expect.

We evaluate out-of-sample rolling forecasting performances and computational times of the proposed methods with the same competing models analyzed in Section 3.4 for comparison. In particular, starting from 1995Q1 (T = 60) to 2019Q2 (T = 120), we fit the corresponding models by means of all available data at time $t-1$ and compute the one step ahead MSFE. Results are summarized in Table 4.

Notice that, on average, the MCMC outperforms all competing alternatives in the forecasting exercise, closely followed by the MAP and ML estimators. We then find the SSSS, CB and CC estimators which perform better than the stacked VAR. Despite its superior forecast accuracy, the MCMC method is more than 140 times slower than the EMVS. This might render the latter preferable in high-dimensional empirical applications, given the relatively little differences between the two in terms of performance.

A key advantage of the EMVS procedure relative to standard MCMC is that its lower computational intensiveness allows for dynamic posterior exploration (Ročková and George, 2018). This consists of holding fixed at a high value the slab hyperparameter, while letting the spike hyperparameter gradually increase along a ladder of increasing values. In our case, dynamic posterior exploration can also be conducted on the "sparsity" parameters θ_k.

We therefore perform cross validation on a grid of ten values between 0.005 and 0.05 with a step of 0.005 for the spike parameter τ_0. Additionally, we let $\beta_k = J_k^{\zeta}$ vary on a grid of ten values of ζ from 0.1 to 1 with step 0.1. By performing dynamic posterior exploration, the average MSFE of the MAP estimator drops to 2.89 $\times 10^{-5}$ (std: 1.69×10^{-5}), while preserving a reasonable amount of computation time of 16.81 seconds (std: 5.86).

Kronecker GFEVD

The lack of a tensor moving average representation requires the vectorized form of the model for dynamic analysis, which in the case of a MAR*(P) reads as

$$\begin{aligned} \mathbf{y}_t &= \sum_{i=1}^{P} c_i (\mathbf{B} \otimes \mathbf{A}) \mathbf{y}_{t-i} + \mathbf{e}_t, \\ \mathbf{e}_t &\sim \mathcal{N}(\mathbf{0}, \boldsymbol{\sigma}_3^2 \otimes \boldsymbol{\Sigma}_2 \otimes \boldsymbol{\Sigma}_1). \end{aligned} \tag{3.24}$$

If all the roots of $|\mathbf{I}_{GN} - \sum_{i=1}^{P} c_i (\mathbf{B} \otimes \mathbf{A}) z^i| = 0$ fall outside the unit circle, the vectorized form of the MAR*(P) admits the following VMA representation

$$\mathbf{y}_t = \sum_{i=0}^{+\infty} \boldsymbol{\Psi}_i \mathbf{e}_{t-i}, \tag{3.25}$$

where $\boldsymbol{\Psi}_i = c_1 (\mathbf{B} \otimes \mathbf{A}) \boldsymbol{\Psi}_{i-1} + ... + c_P (\mathbf{B} \otimes \mathbf{A}) \boldsymbol{\Psi}_{i-P}$, with $\boldsymbol{\Psi}_0 = \mathbf{I}_{GN}$ and $\boldsymbol{\Psi}_i = \mathbf{0}$ for $i < 0$.

Dynamic analysis via GIRF and GFEVD can be thereby carried out easily on a variable by variable basis. In this context, the modification of the

GFEVD by Lanne and Nyberg (2016) obtained from the Generalized Impulse Response Function (GIRF) of Koop et al. (1996) is appealing, as it enjoys the desired property of unit row sum. Let $\mathbf{\Lambda} = \text{diag}(\mathbf{\Sigma}^{-1})$, where $\mathbf{\Sigma} = \mathbf{\Sigma}_3 \otimes \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1$. The GFEVD for the vectorized MAR*(P) has the standard form

$$\theta_H^{ij} = \frac{\sum_{h=0}^{H} \mathbf{e}_i' \mathbf{\Psi}_h \mathbf{\Sigma} \mathbf{\Lambda}_{jj}^{1/2} \mathbf{e}_j}{\sum_{h=0}^{H} \mathbf{e}_j' \mathbf{\Psi}_h \mathbf{\Sigma} \mathbf{\Lambda} \mathbf{\Sigma} \mathbf{\Psi}_h' \mathbf{e}_j}, \tag{3.26}$$

where $\mathbf{e}_i$ is a selection vector. The collection of $\theta_H^{ij} \in \mathbf{\Theta}_H$ is referred to as GFEVD matrix.

Although the MAR model estimates interpretable lower dimensional coefficients via Tucker product, its vectorized representation leads to a significant loss of information on mode specific interactions, which are inherent of the model. Rather than uniquely studying variable by variable impacts, it is convenient to exploit the enhanced interpretability given by the bi-dimensional structure of the model and decompose the GFEVD into two lower dimensional matrices representing the K-mode specific GFEVD. Among all possible decompositions of a GFEVD matrix into its mode-specific counterparts, the most reasonable approach given the form of the model is via a Kronecker decomposition:

$$\mathbf{\Theta}_H^C \otimes \mathbf{\Theta}_H^I \approx \mathbf{\Theta}_H, \tag{3.27}$$

where $\mathbf{\Theta}_H^I \in \mathbb{R}^{G \times G}$ and $\mathbf{\Theta}_H^C \in \mathbb{R}^{N \times N}$ will be, respectively, the country and indicator GFEVD.

We now describe the Kronecker decomposition problem for the GFEVD derived from the MAR model. Recall that, for each forecast horizon, $\mathbf{\Theta}_H$ is a stochastic matrix, having $\mathbf{\Theta}_H \geq \mathbf{0}$ and $\mathbf{\Theta}_H \mathbf{1}_{GN} = \mathbf{1}_{GN}$, where $\mathbf{1}_J$ is a J-dimensional vector of ones. A simple approach to decompose $\mathbf{\Theta}_H$ into $\mathbf{\Theta}_H^C \otimes \mathbf{\Theta}_H^I$, so as to reflect the indicator and country GFEVD structure is to project $\mathbf{\Theta}_H$ onto the space of Kronecker products under the squared Frobenius norm:

$$\min_{\mathbf{\Theta}_H^V, \mathbf{\Theta}_H^C} \quad ||\mathbf{\Theta}_H - \mathbf{\Theta}_H^C \otimes \mathbf{\Theta}_H^I||_F^2, \tag{3.28}$$

which represents a Nearest Kronecker Product (NKP) problem in matrix computation (Van Loan and Pitsianis, 1993; Loan, 2000). This approach is also at the basis of the projection method for MAR estimation outlined by Chen et al. (2021), which can be used to find the starting values of the ML procedure given a VAR estimate.

However, the two resulting matrices minimizing this problem are not guaranteed to be stochastic as well, a necessary condition to constitute GFEVD. Thus, in order to get the best stochastic Kronecker Product (SKP) approximation $\boldsymbol{\Theta}_H^C$ and $\boldsymbol{\Theta}_H^I$, the following constrained nonlinear least squares must be solved:

$$\begin{aligned} \min_{\boldsymbol{\Theta}_H^I, \boldsymbol{\Theta}_H^C} \quad & ||\boldsymbol{\Theta}_H - \boldsymbol{\Theta}_H^C \otimes \boldsymbol{\Theta}_H^I||_F^2, \\ \text{s.t.} \quad & \boldsymbol{\Theta}_H^I \geq \mathbf{0}, \;\; \boldsymbol{\Theta}_H^I \mathbf{1}_G = \mathbf{1}_G, \\ & \boldsymbol{\Theta}_H^C \geq \mathbf{0}, \;\; \boldsymbol{\Theta}_H^C \mathbf{1}_N = \mathbf{1}_N. \end{aligned} \tag{3.29}$$

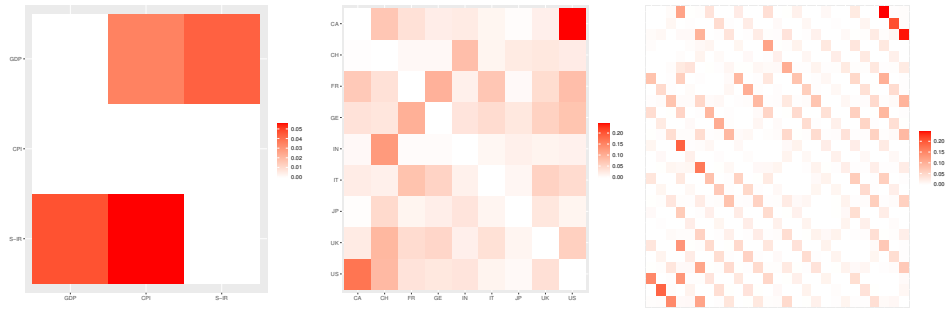
Notice that all the entries in $\boldsymbol{\Theta}_H^C \otimes \boldsymbol{\Theta}_H^I$ are the same as all the entries in $\boldsymbol{\theta}_H^I \boldsymbol{\theta}_H^{C\prime}$, where $\boldsymbol{\theta}_H^I = \text{vec}(\boldsymbol{\Theta}_H^I)$ and $\boldsymbol{\theta}_H^C = \text{vec}(\boldsymbol{\Theta}_H^C)$, i.e. the matrices have the same set of elements, which only differ in their placement. By employing a rearrangement operator $\mathcal{G}(\cdot)$ such that $\mathcal{G}(\mathbf{X} \otimes \mathbf{Y}) = \text{vec}(\mathbf{Y})\text{vec}(\mathbf{X})'$, we can rewrite (3.29) as:

$$\begin{aligned} \min_{\boldsymbol{\theta}_H^I, \boldsymbol{\theta}_H^C} \quad & ||\mathcal{G}(\boldsymbol{\Theta}_H) - \boldsymbol{\theta}_H^I \boldsymbol{\theta}_H^{C\prime}||_F^2 \\ \text{s.t.} \quad & \boldsymbol{\theta}_H^I \geq \mathbf{0}, \;\; \mathbf{R}_I \boldsymbol{\theta}_H^I = \mathbf{1}_G \\ & \boldsymbol{\theta}_H^C \geq \mathbf{0}, \;\; \mathbf{R}_C \boldsymbol{\theta}_H^C = \mathbf{1}_N \end{aligned} \tag{3.30}$$

where $\mathbf{R}_I = [\mathbf{I}_G, ..., \mathbf{I}_G]$ and $\mathbf{R}_C = [\mathbf{I}_N, ..., \mathbf{I}_N]$ are linear equality constraint matrices of dimension $[G \times G^2]$ and $[N \times N^2]$. As it is expressed, the problem in eq. (3.30) can be solved for $\boldsymbol{\theta}_H^I$ and $\boldsymbol{\theta}_H^C$ iteratively via standard constrained minimization routines.

We illustrate the resulting Kronecker decomposition of the estimated GFEVD in Figure 11. For what concerns $\hat{\boldsymbol{\Theta}}_H^I$, a contribution of CPI and GDP to the GFEVD of S-IR is detected. Moreover, the figure shows that a portion of GFEVD in GDP is due to shocks in CPI and S-IR. Notice however

Figure 11: Kronecker decomposition of the GFEVD



(a) $\hat{\boldsymbol{\Theta}}_H^I$ (b) $\hat{\boldsymbol{\Theta}}_H^C$ (c) $\hat{\boldsymbol{\Theta}}_H$

Kronecker decomposition of the estimated GFEVD: indicator GFEVD $\hat{\boldsymbol{\Theta}}_H^I$ (a), country GFEVD $\hat{\boldsymbol{\Theta}}_H^C$ (b) and full GFEVD $\hat{\boldsymbol{\Theta}}_H$ (c). Diagonal elements are omitted.

that overall magnitudes of such cross-variance shares are small if compared to the country and full GFEVD matrices, meaning a weak dependence structure within the indicator dimension.

In the country dimension we find stronger cross-variance shares as reflected by $\hat{\boldsymbol{\Theta}}_H^C$. The largest pairwise contributions are those to Canada arising from shocks in the US economy, and vice versa, though with lower magnitude in the latter case. While both China and US are prone to transmit large portions of GFEVD to the rest of the countries, the former is generally more resilient to shocks in other countries. This with the exception of India, which is one of the largest exporters of China. Results also highlight noticeable cross variance shares across the EU countries, i.e. France, Germany and Italy. The full GFEVD matrix $\hat{\boldsymbol{\Theta}}_H$ reflects instead variable by variable interactions, which seem consistent with the ones obtained through the Kronecker decomposition problem.

3.6 Conclusion

We developed a Bayesian method for variable selection in high-dimensional matrix autoregressive models which reflects and exploits the original matrix structure of data to: a) reduce dimensionality; b) foster interpretability of multidimensional dependency structures. We firstly derive a compact form of the model stemming from the tensor linear regression framework, which facilitates the model estimation. We then outlined two computational methods: a fully Bayesian MCMC algorithm and an EMVS estimation procedure, which foresees the forthcoming need of modeling matrix-valued time series at large scales, while allowing for fast dynamic posterior exploration.

We have numerically investigated the small sample efficiency of the proposed estimators, showing the gain with respect to ML in sparse, high-dimensional settings. We have also numerically explored the comparative estimation, forecasting and computational performances of the proposed estimators relative to key competing alternative models for longitudinal data. The experiment has shown that the estimation and forecasting performances of the Bayesian and MAP estimators are generally superior in sparse high-dimensional settings, with the latter drastically reducing computational intensiveness. The proposed methodology has been applied to a panel of nine world countries' economic indicators, for which we derive a method to decompose the GFEVD into its row and column dimensions, leading to country and indicator GFEVDs.

Our proposed method can be extended in several directions. Simultaneous sparsity both in the autoregressive coefficients and innovation covariance matrices can be introduced. Otherwise, the model can be equipped with different types of priors, e.g. those belonging to the class of global-local priors; see Polson et al. (2012). Furthermore, time variation can be embedded into the model. This also paves the way to the introduction of time varying parameter matrix autoregression with stochastic volatility (Nakajima, 2011) or dynamic sparse factor matrix autoregressions along the lines of Rockova

and McAlinn (2021).

4 A Multidimensional approach to Finance: on the Relationship between Sentiment, Returns and Volatility in the Crypto Market

Based on: Celani, A., and Pagnottoni, P. (2023), "The Multidimensional Relationship between Sentiment, Returns and Volatility". *Available at SSRN.*

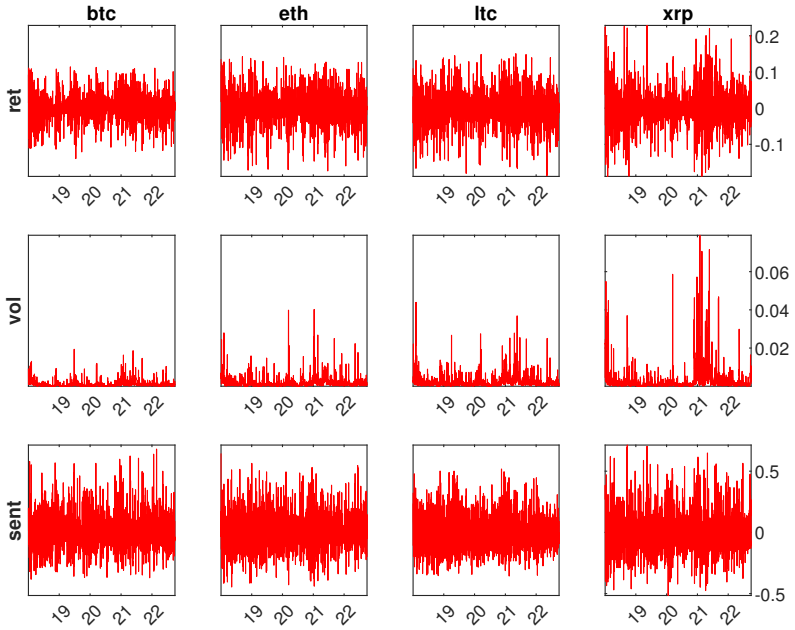
4.1 Introduction

The collection of data at the intersection of two or more classifications naturally constitutes matrix or tensor-valued time series. For instance, the collection of panel data forms matrix-valued observations, whose rows might represent indicators and columns countries. In this context, univariate time series analysis focuses on one element of the matrix at a time. Vector and panel time series analysis deal with the co-movement of one row in the matrix. Modeling each dimension separately annihilates the multidimensional structure of data, and can therefore lead to a significant loss of information or efficiency.

We illustrate the time series of the three indicators for each cryptocurrency in matrix fashion in Figure 12. Some noticeable co-movements are present, especially when considering the time series of the same indicator across the four cryptocurrencies. Indeed, the multivariate time series in matrix form might reflect different types of structures of the time series observations, which would be lost if modeled in a vectorized form.

Against this, we propose an autoregressive model for matrix-valued time series in a frequentist and Bayesian formulation whose most appealing characteristics is to preserve the multidimensional structure of financial time series and admit corresponding interpretations. Its Bayesian formulation produces credible intervals for the estimates might depend upon the prior expectation

Figure 12: Matrix time series of the indicators



Time series of three indicators: Returns, Volatility and Sentiment for the four cryptocurrencies under analysis. Daily data from 01/01/2018 to 09/30/2022.

of the researcher. We then build on Greenwood-Nimmo et al. (2021) and derive a framework for multidimensional connectedness measures, i.e. the connectedness within each dimension of the time series classification. This approach is relatively flexible and opens up the possibility for such models to investigate financial connectedness in depth and breadth. Our method enhances interpretability of multidimensional relationship structures, while reducing dimensionality, at relatively no cost in terms of predictive accuracy.

We apply our model to study the interrelated dynamics of a set of cryptocurrency (i.e. Bitcoin, Ethereum, Litecoin and Ripple) returns, volatility and sentiment, as proxied by Google search volumes. Our empirical applica-

tion shows that spillovers in the cryptocurrency dimension are considerably higher than those in the indicator dimension; thereby interconnectedness across return, volatility and sentiment indicators for the same cryptocurrency is generally weak compared to that measured among the same indicators for different cryptocurrencies. As far as the indicator dimension is concerned, sentiment seems to be the most influent variable in the spillover transmission mechanism, though with a relatively low magnitude, whereas volatility the least influent and most dependent. This is in accord with the extant literature on the explanatory and predictive power of investor sentiment, being it news or social media sentiment, or search volumes, on returns and volatility. The cryptocurrency dimension shows instead no clear ranking in terms of influence across cryptocurrencies, apart from the evidence that Ripple emerges as a net shock receiver. More interestingly, we observe that during a high-volatile market phase dependence drastically drops, meaning the degree of synchronicity of indicators belonging to the same cryptocurrency drastically falls. Finally, we conduct a comparative forecasting performance experiment which confirms the satisfactory forecast accuracy of the proposed model and its superior performances in high-dimensional settings.

Our study is related to the literature on forecasting cryptocurrency market movements - see Catania et al. (2019); Atsalakis et al. (2019); Chen et al. (2021); Gradojevic et al. (2021), but takes a further step towards the explanation of multidimensional relationships across cryptocurrency price, volatility and investor attention.

The remainder of this chapter is structured as follows. Section 4.2 presents the theoretical financial framework, its literature review and the empirical motivation of the study. Section 4.3 introduces autoregressive models for matrix-valued time series and illustrates the proposed methodology. In Section 4.4 we conduct the empirical application to cryptocurrency returns, volatility and sentiment. Section 4.6 concludes.

4.2 Financial theory, literature and motivation

The financial literature has generally agreed on the fact that asset prices are driven by the actions of two types of investors: arbitrageurs and noise traders - see Shleifer and Summers (1990); Harris and Raviv (1993); Stambaugh et al. (2012); Asness et al. (2013). While arbitrageurs rely on fundamentals to trade, and thereby make efforts to bring prices in harmony with the "true" fundamental value, noise traders tend to trade on noise, i.e. pseudo-signals, or other kind of trading models which alter demand and, as a consequence, prices. This is corroborated by a number of empirical studies which demonstrate the effects of attention and news on the buying behavior of individual and institutional investors - see, for instance, Grullon et al. (2004); Barber and Odean (2008); Fang and Peress (2009). De Long et al. (1990) develop a model explaining how noise trader risk is priced in the financial markets. Indeed, while in the long term prices will tend to revert to the fundamental values, this might come with fluctuations and after a considerable amount of time, to the detriment of arbitrageurs in such cases. At the individual asset level, their model generates predictions regarding the relationship between sentiment and price volatility, concluding that greater noise trading is related to higher price volatility. Additionally, sentiment will exert effects on returns through its impact on volatility. As a consequence, it is expected to find an association between sentiment indicators and returns and volatility when sentiment is the signal driving noise trading.

Although noise traders might use different models whose trade outcomes counterbalance, the phenomenon of aggregate demand shift due to correlated strategies is a documented empirical fact in finance. Using survey data, Schmeling (2007) discovers that stock market returns are predicted by the sentiment of individual investors. In a similar way, Barber et al. (2009) employ brokerage data and find that individual investors frequently buy the same stocks concurrently, which results in shifting prices in an upward direction. These studies essentially show that attempts by arbitrageurs to bring

prices back in accord with fundamentals are not always successful, and that changes in the demand for securities that are unrelated to fundamentals may be evident and last over time. Since the supply curve for stocks is, at least in the short run, inelastic, the purchasing pressure on stocks put forward by increased search behavior should exert a buying shock which translates into a significant and rapid increase in stock prices. In other words, before arbitrageurs might correct any mispricing, an abnormal return should follow any consistent buying shock due to an increased search activity, which makes financial markets a particularly attractive environment in which to evaluate the impact of search activity.

The relationships across returns and sentiment, volatility and sentiment, and the joint relations among the three has been explored in a large number of studies; notwithstanding this, mixed results have been found.

The extant literature has investigated the role of sentiment and attention indicators to predict financial market returns. Many of them conclude that public sentiment, as measured by indicators built upon social media attention and internet search volumes, can predict stock market returns - see, among others, Neal and Wheatley (1998); Da et al. (2011); Joseph et al. (2011); Da et al. (2015); Sun et al. (2016); Renault (2017); Gu and Kurov (2020). Nevertheless, not all studies which examined at the relationship between sentiment and returns came to the same results, and some even suggest that the causality direction is from returns to sentiment, rather than vice-versa. Brown and Cliff (2004) study the link between sentiment and equity returns using a wide variety of sentiment indicators and discover significantly stronger evidence that sentiment is influenced by returns, a conclusion which is in line also with Solt and Statman (1988). According to Wang et al. (2006), returns (and volatility), not the other way around, are the indicators influencing the majority of their investigated sentiment measurements. Among others, Fisher and Statman (2000) instead find that there might be a significant causal relationship in both directions between equity returns and sentiment.

Similarly, many papers have studied the relationship between sentiment and volatility. A vast majority conclude that investor sentiment, being it social media, news sentiment or search volumes, are useful predictors of the stock market volatility - see Antweiler and Frank (2004); Ho et al. (2013); Aouadi et al. (2013); Hamid and Heiden (2015); Dimpfl and Jank (2016); Siganos et al. (2017); Behrendt and Schmidt (2018); Liang et al. (2020); Chen et al. (2021); Caporin and Poli (2022). In contrast, a few exceptions find the opposite. Wang et al. (2006) find that the analyzed sentiment indicators are caused by volatility (and returns) rather than the other way around. They provide evidence on the fact that lagged returns cause volatility, and that sentiment variables have extremely limited forecasting power when returns are included as a forecasting variable.

When it comes to cryptocurrencies, the actual drivers of their prices and sources of their extreme volatility are still unclear and widely discussed by scholars. While part of the research has tested whether these might be associated with the dimension of the user base (Van Vliet, 2018), investor expectations of future price movements (Koutmos and Payne, 2021) and cost of mining (Hayes, 2019; Kristoufek, 2020), it turns out that a component due to irrational bubbles and investor behaviour is reasonably part of the price formation process - see, e.g., Cheah and Fry (2015); Anyfantaki et al. (2021); Dimpfl and Peter (2021); El Montasser et al. (2022); Koch and Dimpfl (2023). Investor behaviour is still mostly irrational, given the lack of fundamentals and the technical skills needed for a complete understanding of the underlying blockchain technology of each of the securities, which makes it extremely difficult to value them from a rational viewpoint.

The importance of investor attention in explaining cryptocurrency returns and volatility has been documented by several studies so far, although not all of the literature agrees on the dominant directions of mutual influences. On the one hand, Kapar and Olmo (2021) conclude that the efficient price of cryptocurrencies behaves idiosyncratically, specifically during turmoils, and can be only rationalised by individuals' Google search volumes. Bleher and

Dimpfl (2019), in evaluating whether Google search volume is a good predictor of cryptocurrency returns and volatility, find that returns are not predictable via internet queries, whereas volatilities are. On the other hand, the results of Neto (2021) shed light on a unidirectional causality effect from cryptocurrency returns to Google search volumes in the first place, which turns to be bidirectional as the time delay increases. A similar conclusion is reached by Lin (2021), who reports that cryptocurrency returns exert a significant impact on future google search volumes, and not viceversa. Using Twitter-based attention measures, Kraaijeveld and De Smedt (2020) show a significant predictive power of investor attention for different cryptocurrency returns, as also corroborated by Shen et al. (2019), who find that such attention measures can predict both Bitcoin volume and volatility. A few studies employ both Google search volumes and Twitter-based attention measures, such as Philippas et al. (2019) and Koch and Dimpfl (2023). The first one suggests that prices are partially driven by attention measures, and that the relationship becomes stronger in periods of high uncertainty, whereas the second one concludes that rising attention Granger-causes increasing price synchronicity of cryptocurrencies.

To sum up, the literature generally agrees that sentiment might be a good predictor for financial market volatility, though the relationship between the two could be influenced by returns (Wang et al., 2006). Moreover, when looking at the cryptocurrency market, the lack of fundamental values paves the way for potentially stronger relations across sentiment, returns and volatility, which, at the moment, remain highly debated in terms of their magnitude and, more importantly, direction of influence. This is why, in our empirical application, we jointly model the dynamics of sentiment, returns and volatility, so to unveil the direction of the relationships existing among the three.

4.3 The Matrix Autoregressive Model

The key feature of the Matrix Autoregressive (MAR) model is to treat time series observations lying at the intersection of two dimensions as a matrix, and then model them as an autoregressive process. In our case, the first dimension refers to a number $g = 1, ..., G$ of indicators useful to describe the behavior of different assets $n = 1, ..., N$, the second dimension of interest. In particular, let $r_{n,t}$, $v_{n,t}$, and $s_{n,t}$ be, respectively, the time $t = 1, ..., T$ n-th asset return, volatility and sentiment, and define $\mathbf{y}_{n,t} = [r_{n,t}, v_{n,t}, s_{n,t}]^{'}$ as the G-dimensional vector of indicators belonging to the generic asset n observed at time t. By horizontally stacking all the $\mathbf{y}_{n,t}$, we obtain the $G \times N$ matrix of dependent variables $\mathbf{Y}_t = [\mathbf{y}_{1,t}, ..., \mathbf{y}_{N,t}]$. A MAR model of order P is then defined as:

$$\mathbf{Y}_t = \mathbf{M} + \mathbf{A}_1 \mathbf{Y}_{t-1} \mathbf{B}_1^{'} + ... + \mathbf{A}_P \mathbf{Y}_{t-P} \mathbf{B}_P^{'} + \mathbf{E}_t \tag{4.1}$$

$$\mathbf{E}_t \sim \mathcal{MN}(\mathbf{0}, \mathbf{\Sigma}_1, \mathbf{\Sigma}_2) \tag{4.2}$$

where $\mathbf{M}$ is a $G \times N$ matrix of constants, $\mathbf{A}_p$ and $\mathbf{B}_p$ for $p = 1, ..., P$ are left autoregressive matrices of dimension $G \times G$ and $N \times N$, respectively. $\mathbf{E}_t$ is a matrix error term that follows a matrix-variate Normal distribution (Gupta and D.K., 1999) with positive definite left and right covariance matrices $\mathbf{\Sigma}_1$ and $\mathbf{\Sigma}_2$ of the same dimension of $\mathbf{A}_p$ and $\mathbf{B}_p$.

By applying the vectorization operator to equation (4.1), it cleary emerges that every MAR has an equivalent VAR representation. Let $\mathbf{y}_{t-p} = \text{vec}(\mathbf{Y}_{t-p})$ for $p = 0, ..., P$ and $\mathbf{e}_t = \text{vec}(\mathbf{E}_t)$, then we obtain:

$$\mathbf{y}_t = \mathbf{m} + (\mathbf{B}_1 \otimes \mathbf{A}_1)\mathbf{y}_{t-1} + ... + (\mathbf{B}_P \otimes \mathbf{A}_P)\mathbf{y}_{t-P} + \mathbf{e}_t \tag{4.3}$$

$$\mathbf{e}_t \sim \mathcal{N}(\mathbf{0}, \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1) \tag{4.4}$$

n where $\mathbf{m} = \text{vec}(\mathbf{M})$ and $\otimes$ is the Kronecker product.

As a direct consequence of the VAR representation of the MAR, the MAR(P) in equation (4.1) is stationary if all the roots of $|\mathbf{I}_{GN} - (\mathbf{B}_1 \otimes \mathbf{A}_1)z +$

$... + (\mathbf{B}_P \otimes \mathbf{A}_P)z^P| = 0$ fall outside the unit circle. Under this assumption, the model in equation (4.1) can be rewritten as the infinite moving average representation,

$$\mathbf{y}_t = \mathbf{e}_t + \mathbf{\Psi}_1 \mathbf{e}_{t-1} + ... + \mathbf{\Psi}_\infty \mathbf{e}_{t-\infty} \tag{4.5}$$

where the $GN \times GN$ coefficient matrices $\mathbf{\Psi}_k$ can be obtained via the following recursive relations

$$\mathbf{\Psi}_k = (\mathbf{B}_1 \otimes \mathbf{A}_1)\mathbf{\Psi}_{k-1} + ... + (\mathbf{B}_P \otimes \mathbf{A}_P)\mathbf{\Psi}_{k-P} \tag{4.6}$$

where $\mathbf{\Psi}_1 = \mathbf{I}_{GN}$ and $\mathbf{\Psi}_k = \mathbf{0}$ for $k < 1$.

Intepretation

Given the one-to-one relationship between equations (4.1) and (4.3), it is evident that a MAR is equivalent to a VAR with specific restrictions on the coefficient matrices, generated by the Kronecker product. As such, dimension-specific interactions arise from this decomposition, and the MAR approach emerges as an effective dimensionality reduction technique. The nature of the restriction, induced just by treating temporal observations as matrices, has the non-negligible benefit of being a hyperparameter-free dimensionality reduction technique. In fact, despite the great flexibility of the multiple dimensionality reduction techniques proposed in the literature so far, none of them prevents the researcher from the need to rely on usually hard apriori assumptions (Bai and Ng, 2002; Forni et al., 2005; George et al., 2008; Bańbura et al., 2010; Koop and Korobilis, 2013; Kock and Callot, 2015; Ahelegbey et al., 2022). As anticipated, this generally done by tuning a number of hyperparameters, a cumbersome task in high-dimensional setting.

To ease interpretation, assume $\mathbf{M} = \mathbf{0}$ and $P = 1$. Then the conditional mean for the j-th asset return can be rewritten as:

$$\mathrm{E}(r_{j,t}|\mathbf{Y}_{t-1}) = a_{11} \sum_{n=1}^{N} b_{jn} r_{n,t-1} + a_{12} \sum_{n=1}^{N} b_{jn} v_{n,t-1} + a_{13} \sum_{n=1}^{N} b_{jn} s_{n,t-1}, \tag{4.7}$$

$a_{ij} \in \mathbf{A}_1$ and $b_{ij} \in \mathbf{B}_1$. Equation (4.7) implies that the MAR framework the conditional mean for the j-th asset return is a function of two nested linear combinations. The first one expresses each indicator as a linear combination of, respectively, the past returns, volatilities and volumes, with constant coefficients given by the j-th row of $\mathbf{B}$. These elements can be viewed as *market adjusted indicators*, representing an average market movement for each indicator. The diagonal elements embed the effect of the indicator dimension on its past values, whereas the off diagonal ones reflect the impact of the other indicators. The second one is a linear combination of the market adjusted indicators with constant coefficients given by the i-th row of $\mathbf{A}$. Such elements can be viewed as *variable adjusted indicators*, which reflect the average asset co-movement in their market adjusted indicators. The diagonal elements embed the effect of the asset dimension on its past values, whereas the off diagonal ones reflect the impact of the other assets.

Estimation

In this section we present the inference procedure for the model in equation (4.1). For this purpose, it is convenient to rewrite the MAR model more compactly for each t. In particular, let

$$\mathbf{X}_{t-1} = \begin{bmatrix} \mathbf{Y}_{t-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{Y}_{t-2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{Y}_{t-P} \end{bmatrix} \tag{4.8}$$

be the $GP \times NP$ matrix of endogenous regressors. Then equation (4.1) rewrites as:

$$\mathbf{Y}_t = \mathbf{M} + \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}^{'} + \mathbf{E}_t \tag{4.9}$$

where $\mathbf{A} = [\mathbf{A}_1, ..., \mathbf{A}_P]$ and $\mathbf{B} = [\mathbf{B}_1, ..., \mathbf{B}_P]$ are the stacked lagged left and right coefficient matrices, respectively. The log-likelihood of the model up to

a constant is then given by:

$$\begin{aligned}\mathcal{L} = &- N(T-P)\log|\boldsymbol{\Sigma}_1| - G(T-P)\log|\boldsymbol{\Sigma}_2| \\ &- \sum_t \operatorname{tr}\Big[\boldsymbol{\Sigma}_1^{-1}(\mathbf{Y}_t - \mathbf{M} - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')\boldsymbol{\Sigma}_2^{-1}(\mathbf{Y}_t - \mathbf{M} - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')\Big]\end{aligned} \tag{4.10}$$

The estimation of $\mathbf{M}$, $\mathbf{A}$, $\mathbf{B}$, $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$ can be obtained by employing an alternating ML algorithm. At each step, we iteratively update one matrix while holding the other fixed, and vice-versa. The steps for the estimation of a MAR(P) can be carried out by generalizing the procedure for a MAR(1) outlined in Chen et al. (2021). The steps are given by:

$$\begin{aligned}\mathbf{M} =& \frac{1}{T-P}\sum_t(\mathbf{Y}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}') \\ \mathbf{A} =& \left(\sum_t(\mathbf{Y}_t - \mathbf{M})\boldsymbol{\Sigma}_2^{-1}\mathbf{B}\mathbf{X}_{t-1}'\right)\left(\sum_t \mathbf{X}_{t-1}\mathbf{B}'\boldsymbol{\Sigma}_2^{-1}\mathbf{B}\mathbf{X}_{t-1}'\right)^{-1} \\ \mathbf{B} =& \left(\sum_t(\mathbf{Y}_t - \mathbf{M})'\boldsymbol{\Sigma}_1^{-1}\mathbf{A}\mathbf{X}_{t-1}\right)\left(\sum_t \mathbf{X}_{t-1}'\mathbf{A}'\boldsymbol{\Sigma}_1^{-1}\mathbf{A}\mathbf{X}_{t-1}\right)^{-1} \\ \boldsymbol{\Sigma}_1 =& \frac{\sum_t(\mathbf{Y}_t - \mathbf{M} - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')'\boldsymbol{\Sigma}_1^{-1}(\mathbf{Y}_t - \mathbf{M} - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')}{N(T-P)} \\ \boldsymbol{\Sigma}_2 =& \frac{\sum_t(\mathbf{Y}_t - \mathbf{M} - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')\boldsymbol{\Sigma}_2^{-1}(\mathbf{Y}_t - \mathbf{M} - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')'}{G(T-P)}\end{aligned} \tag{4.11}$$

Prior formulation

To ease the exposition, we consider a MAR(P) with no constant. Posterior inference for the model in equation (4.1) under an Independent Normal Wishart framework is derived by assuming a Normal prior for the vectorized version of $\mathbf{A}$ and $\mathbf{B}$:

$$\begin{aligned}\pi(\boldsymbol{\alpha}) &\sim \mathcal{N}(\boldsymbol{\alpha}_0, \Omega_\alpha) \\ \pi(\boldsymbol{\beta}) &\sim \mathcal{N}(\boldsymbol{\beta}_0, \Omega_\beta)\end{aligned} \tag{4.12}$$

where $\boldsymbol{\alpha} = \text{vec}(\mathbf{A})$ and $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$. The two covariance matrices are however not separately identifiable from the likelihood, hence prior restrictions have to be used to achieve identification (for a deeper explanation see (Hoff, 2015; Billio et al., 2022) among others). For the pair of covariance matrices $\boldsymbol{\Sigma}_1$ and $\boldsymbol{\Sigma}_2$, we assume the following hierarchical prior distribution:

$$\begin{aligned} \pi(\gamma) &\sim \mathcal{G}a(a_0, b_0) \\ \pi(\boldsymbol{\Sigma}_1|\gamma) &\sim \mathcal{IW}(\gamma S_1, \nu_1) \\ \pi(\boldsymbol{\Sigma}_2|\gamma) &\sim \mathcal{IW}(\gamma S_2, \nu_2) \end{aligned} \tag{4.13}$$

Posterior approximation

By combining the log-likelihood in equation (4.10) with prior assumptions we can analytically derive the full conditional distributions of $\mathbf{A}, \mathbf{B}, \boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2, \gamma$. Since the joint posterior distribution is not tractable, we implement a Gibbs sampler that iterates over the following steps:

- Draw $\boldsymbol{\Sigma}_1$ from $\mathcal{IW}(\overline{S}_1, \nu_1 + G)$, where $\overline{S}_1 = \gamma S_1 + \sum_t (\mathbf{Y}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')'\boldsymbol{\Sigma}_1^{-1}(\mathbf{Y}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')$
- Draw $\boldsymbol{\alpha}$ from $\mathcal{N}(\overline{\boldsymbol{\alpha}}, \overline{\Omega}_\alpha)$, where $\overline{\Omega}_\alpha = [\Omega_\alpha^{-1} + (\sum_t \mathbf{X}_{t-1}\mathbf{B}'\boldsymbol{\Sigma}_2^{-1}\mathbf{B}\mathbf{X}_{t-1}') \otimes \boldsymbol{\Sigma}_1^{-1}]^{-1}$ and $\overline{\boldsymbol{\alpha}} = \overline{\Omega}_\alpha[\Omega_\alpha\boldsymbol{\alpha}_0 + \text{vec}((\sum_t \mathbf{Y}_t\boldsymbol{\Sigma}_2^{-1}\mathbf{B}\mathbf{X}_{t-1}')(\sum_t \mathbf{X}_{t-1}\mathbf{B}'\boldsymbol{\Sigma}_2^{-1}\mathbf{B}\mathbf{X}_{t-1}')^{-1})]$
- Draw $\boldsymbol{\Sigma}_2$ from $\mathcal{IW}(\overline{S}_2, \nu_2 + N)$, where $\overline{S}_2 = \gamma S_2 + \sum_t (\mathbf{Y}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')\boldsymbol{\Sigma}_2^{-1}(\mathbf{Y}_t - \mathbf{A}\mathbf{X}_{t-1}\mathbf{B}')'$
- Draw $\boldsymbol{\beta}$ from $\mathcal{N}(\overline{\boldsymbol{\beta}}, \overline{\Omega}_\beta)$, where $\overline{\Omega}_\beta = [\Omega_\beta^{-1} + (\sum_t \mathbf{X}_{t-1}'\mathbf{A}'\boldsymbol{\Sigma}_1^{-1}\mathbf{A}\mathbf{X}_{t-1}) \otimes \boldsymbol{\Sigma}_2^{-1}]^{-1}$ and $\overline{\boldsymbol{\beta}} = \overline{\Omega}_\beta[\Omega_\beta\boldsymbol{\beta}_0 + \text{vec}((\sum_t \mathbf{Y}_t'\boldsymbol{\Sigma}_1^{-1}\mathbf{A}\mathbf{X}_{t-1})(\sum_t \mathbf{X}_{t-1}'\mathbf{A}'\boldsymbol{\Sigma}_1^{-1}\mathbf{A}\mathbf{X}_{t-1})^{-1})]$
- Draw γ from $\pi(\gamma|\boldsymbol{\Sigma}_1, \boldsymbol{\Sigma}_2) \sim \mathcal{G}a\Big(a_0 + \frac{1}{2}\Big[\nu_1 G + \nu_2 N\Big], b_0 + \frac{1}{2}\Big[\text{tr}(S_1\boldsymbol{\Sigma}_1^{-1}) + \text{tr}(S_2\boldsymbol{\Sigma}_2^{-1})\Big]\Big)$

We refer the reader to Celani and Pagnottoni (2022) for further details on the derivation of the posterior distributions of the parameters.

Multidimensional connectedness measures

The connectedness framework based upon Generalized Forecast Error Variance Decomposition (GFEVD) developed by Diebold and Yilmaz (2012, 2014) is well suited to the analysis of relatively small-sized VAR models. For a sufficiently large number of variables, it becomes hard to interpret the relationships arising from the elements on a disaggregated basis. While this issue could be at least partially recovered via dimensionality reduction, the methodological framework is not well suited to the study of potentially large, multidimensional models, as its levels of aggregation of the information are essentially two: a) no aggregation, with bilateral connectedness measures being either determined individually, or on a variable-by-variable basis; or b) complete aggregation, where bilateral connectedness measures are aggregated into single spillover indices. Therefore, the framework as such does not admit intermediate levels of aggregation, such as connectedness across financial assets in aggregate rather than among individual variables, or aggregate linkages across variables in aggregate rather than among individual assets. We propose, in line with Greenwood-Nimmo et al. (2021), a simple approach to tackle both issues based on re-normalisation and block aggregation of the connectedness matrix. Such method can yield generalized connectedness measures embedding multidimensional relationship structures, preserving the original matrix structure of the data.

Dynamic analysis via GIRF and GFEVD can be carried out easily on a variable by variable basis. In this context, the modification of the GFEVD by Lanne and Nyberg (2016) obtained from the Generalized Impulse Response Function (GIRF) of Koop et al. (1996) is appealing, as it boasts the desired property of unit row sum. Let $\mathbf{\Lambda} = \text{diag}(\mathbf{\Sigma}^{-1})$, where $\mathbf{\Sigma} = \mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1$. The GFEVD for the vectorized MAR(P) has the standard form

$$GFEVD(y_{it}, e_{jt}, H) = \theta_{ij}^{H} = \frac{\sum_{h=0}^{H} \mathbf{e}_i^{'} \mathbf{\Psi}_h \mathbf{\Sigma} \mathbf{\Lambda}_{jj}^{1/2} \mathbf{e}_j}{\sum_{h=0}^{H} \mathbf{e}_j^{'} \mathbf{\Psi}_h \mathbf{\Sigma} \mathbf{\Lambda} \mathbf{\Sigma} \mathbf{\Psi}_h^{'} \mathbf{e}_j} \tag{4.14}$$

where $\mathbf{e}_i$ is a selection vector with all zeros and one in the i-th element. The

collection of $\theta_{ij}^H \in \mathbf{\Theta}^H$ is referred to as the GFEVD (connectedness) matrix.

Nevertheless, the dimension specific influences θ^H among indicators and assets cannot be easily recovered from $\mathbf{\Theta}^H$. This emerges as a consequence of conducting dynamic analyses through the vectorized form of the MAR. Following Greenwood-Nimmo et al. (2021), we outline a simple yet effective method to overcome this issue, flexible enough to be accommodated also in a MAR framework. In particular, we design an aggregation scheme to disentangle the effect of indicators and assets in a multi-variable, multi-asset context. Consider the sub-GFEVD $\mathbf{\Theta}_I^H$ to be the $G \times G$ matrix embedding indicator-specific influences. Further define $\mathcal{I}_g = \{g, g+G, ..., g+G(N-1)\}$ as the set of positions of the indicator g in $\mathbf{\Theta}^H$. Then the generic element of $\mathbf{\Theta}_I^H$ is simply defined as

$$\theta_{I,g_1g_2}^H = \sum_{i \in \mathcal{I}_{g_1}} \sum_{j \in \mathcal{I}_{g_2}} \theta_{ij}^H \tag{4.15}$$

Analogously, let $\mathbf{\Theta}_C^H$ pe the $N \times N$ sub-GFEVD embedding crypto influences. In this case, the set of positions where crypto n appears in $\mathbf{\Theta}^H$ is $\mathcal{I}_n = \{(n-1)*G+1, (n-1)*G+2, ..., n*G\}$. Consequently, its generic element is:

$$\theta_{C,n_1n_2}^H = \sum_{i \in \mathcal{I}_{n_1}} \sum_{j \in \mathcal{I}_{n_2}} \theta_{ij}^H \tag{4.16}$$

Given that this aggregation scheme does not ensure that the two sub-GEFVDs have a unit row sum, a normalization scheme such as those employed in Diebold and Yilmaz (2012) and Greenwood-Nimmo et al. (2021) is needed. Therefore, the collection of the row normalized $\theta_{I,ij}^H \in \mathbf{\Theta}_I^H$ and $\theta_{C,ij}^H \in \mathbf{\Theta}_C^H$ are referred to the indicator and asset sub-GFEVD matrices, respectively.

For each of the two sub-GFEVDs, we define the total H step ahead spillover *from*, *to* and *within* as follows:

$$\mathcal{F}_{i \leftarrow \bullet}^H = \sum_{j=1, j \neq i}^{K} \theta_{ij}^H, \qquad \mathcal{T}_{\bullet \leftarrow i}^H = \sum_{j=1, j \neq i}^{K} \theta_{ji}^H, \qquad \mathcal{W}_{i \leftarrow i}^H = \theta_{ii}^H \tag{4.17}$$

In a similar way, we define the aggregate spillover index as:

$$\mathcal{S}^H = \sum_{i=1}^{K} \mathcal{F}^H_{i\leftarrow\bullet} \equiv \sum_{i=1}^{K} \mathcal{T}^H_{\bullet\leftarrow i} \tag{4.18}$$

Finally, we employ a pair of indices to measure: a) the level of depenence of the i-th indicator (asset) on external conditions; and b) to what extent the i-th indicator (asset) influences the system. To answer the first question, we define the *dependence* index:

$$\mathcal{O}^H_i = \frac{\mathcal{F}^H_{i\leftarrow\bullet}}{\mathcal{F}^H_{i\leftarrow\bullet} + \mathcal{W}^H_{i\leftarrow i}} \tag{4.19}$$

where $0 \leq \mathcal{O}^H_i \leq 1$. Similarly, we define the *influence* index as:

$$\mathcal{I}^H_i = \frac{\mathcal{T}^H_{\bullet\leftarrow i} - \mathcal{F}^H_{i\leftarrow\bullet}}{\mathcal{T}^H_{\bullet\leftarrow i} + \mathcal{F}^H_{i\leftarrow\bullet}} \tag{4.20}$$

where $-1 \leq \mathcal{I}^H_i \leq 1$.

4.4 Real application

We demonstrate the properties of our proposed method through a real application to study the multidimensional interconnectedness of a set of cryptocurrency prices, volatility and Google search volumes. Firstly, we provide a full sample (static) analysis to show the properties of the model. Secondly, we study the dynamic evolution of spillovers across the two categories of interest: indicator and crypto (asset) dimensions.

We select four major cryptocurrencies, i.e. Bitcoin (btc), Ethereum (eth), Litecoin (ltc), Ripple (xrp) and three indicators, namely continuous price returns (ret), the realized volatility (vol) and Google query index (sent). The sample period ranges from 01/01/2018 to 09/30/2022, for a total of $T = 1733$ observations. In our application, volatility is measured following Garman and Klass (1980) and Alizadeh et al. (2002). In other words, we use

daily high, low, opening and closing prices obtained from underlying daily high/low/open/close data to estimate return volatility:

$$\begin{aligned} v_{n,t} =& 0.511(H_{n,t} - L_{n,t})^2 - 0.019[(C_{n,t} - O_{n,t})(H_{n,t} + L_{n,t} - 2O_{n,t}) \\ & - 2(H_{n,t} - O_{n,t})(L_{n,t} - O_{n,t})] - 0.383(C_{n,t} - O_{n,t})^2 \end{aligned} \tag{4.21}$$

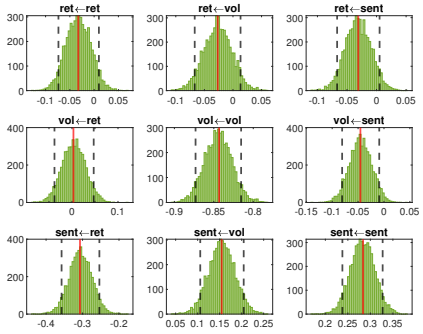
where H is the daily high, L is the low, O is the open and C is the close (all in natural logarithms).

For the purpose of illustrating our method, we estimate a Bayesian MAR(1) which yields the full conditional posterior distribution of the parameters of interest. As far as the prior setting, we use diffuse but proper priors. In particular, we set the following hyperparameters: $\boldsymbol{\alpha}_0 = \mathbf{0}$, $\boldsymbol{\beta}_0 = \mathbf{0}$, $\Omega_\alpha = 100\mathbf{I}_{G^2}$, $\Omega_\beta = 100\mathbf{I}_{N^2}$, $a_0 = b_0 = 0.1$, $S_1 = \mathbf{I}_G/G$, $S_2 = \mathbf{I}_N/N$, $\nu_1 = G+2$, $\nu_2 = N+2$.

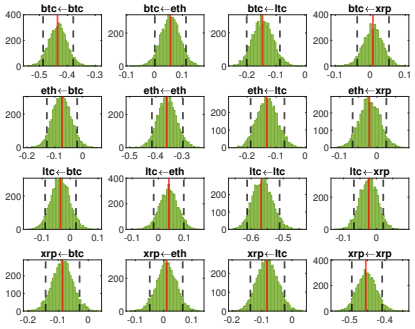
We illustrate in Figure 13 the posterior distribution of the left ($\hat{\mathbf{A}}$) and right ($\hat{\mathbf{B}}$) first order coefficient matrices, and of the left ($\hat{\boldsymbol{\Sigma}}_1$) and right ($\hat{\boldsymbol{\Sigma}}_2$) variance-covariance matrices of innovations, along with their estimated posterior median and 95% credible intervals. We evaluate in Appendix A.3 the convergence of the MCMC algorithm via the diagnostic proposed by Raftery and Lewis (1970), and document that the convergence of the Gibbs sampler is achieved. In particular, we rely on their procedure to select the best combination of initial burn-in periods and thinning factor after having run the MC for a total of interations suggested by the methodology of Gong and Flegal (2016); Vats et al. (2019). We are motivated to employ both the methodologies because we observed instabilities of the MC in the rolling exercise. The reduction in the autocorrelation of the MC, thanks to the cleaning procedure of Raftery and Lewis (1970) is acceptable to carry out inference.

It is evident from the figure that diagonal elements of the parameter matrices are generally of larger magnitude if compared to off-diagonal ones. As far as the left coefficient matrix, volatility seems to be the indicator whose past values impact in a larger way current ones, followed by sentiment

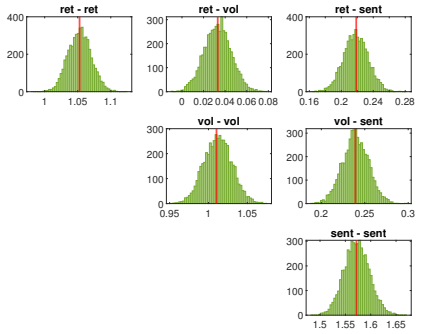
Figure 13: Posterior distribution of coefficient matrices



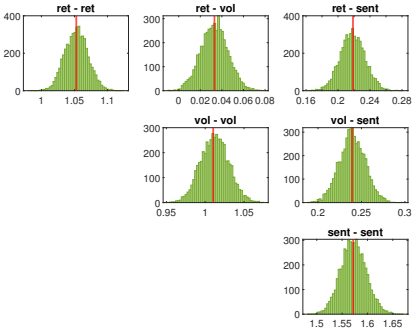
(a) $\hat{\mathbf{A}}$



(b) $\hat{\mathbf{B}}$



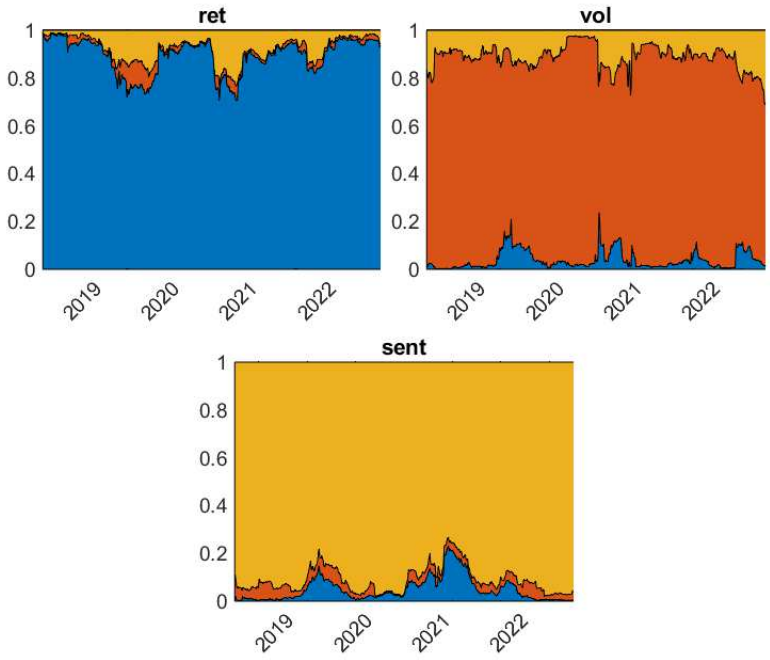
(c) $\hat{\mathbf{\Sigma}}_1$



(d) $\hat{\mathbf{\Sigma}}_2$

Estimated posterior distribution of the left and right first order coefficient matrices (first row) and of the left and right variance-covariance matrix of innovations (second row). Red vertical lines correspond to ML estimates. Dotted black vertical lines delimit the 95% credible intervals.

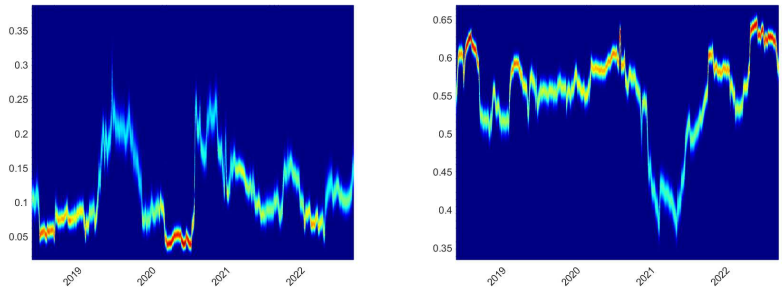
Figure 14: Posterior distribution of the dynamic indicators GFEVD.



and returns, allegedly given by the fact that volatility is a naturally more persistent stochastic process. Further, the credible intervals of the latter include zero, confirming the inherent characteristic of non-predictability of financial returns. Cross-impacts highlight the weak explanatory power of returns on volatility, and viceversa, which is corroborated by the just as low contemporaneous covariance among the two. On the other hand, a significant relationship between sentiment and volatility appears in both directions, and a significant one from sentiment to returns, indicators for which also the contemporaneous covariance is of sizeable magnitude.

As far as the right coefficient matrix is concerned, Litecoin presents the highest autoregressive coefficient, followed by Ripple, Bitcoin and Ethereum. Cross-influences of Bitcoin and Litecoin are the ones of most considerable magnitudes, whereas Ethereum and, especially, Ripple show non-significant

Figure 15: Posterior of the spillover indices



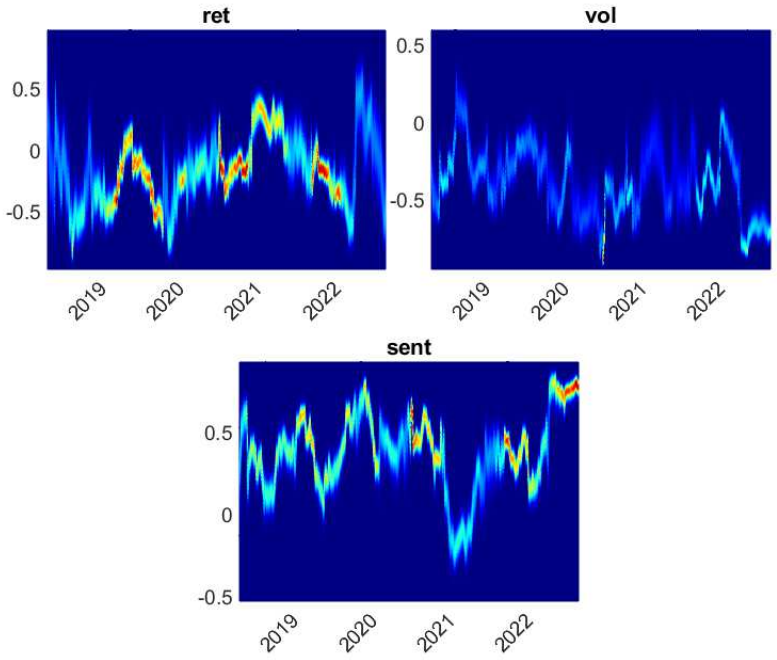
(a) (b)

Posterior distribution of the spillover index for the indicator (panel a) and crypto (panel b) dimension.

lagged impacts on others. Further notice that contemporaneous variances are almost of comparable magnitudes relative to covariances, highlighting a higher degree of contemporaneous correlation in the crypto dimension, rather than in the indicator one. We have further performed dynamic analysis via generalized impulse response functions (GIRFs) arising from a global shock in each indicator; results associated to a positive 1-standard-deviation shock are reported in Figure 34 in Appendix.

Full-sample connectedness provides a static, average representation of market spillovers. However, several changes and noticeable market movements took place during our sample period. Against the background of cryptocurrency market evolution and turmoil, it seems at least partially inadequate to model connectedness in a static fashion. It seems unfeasible for any single fixed-parameter model to successfully apply over the entire sample. To cope with this issue, we estimate dynamic connectedness measures based on rolling samples, so to gauge the nature and magnitude of variations in spillovers over time by analyzing the corresponding time series of spillover measures. This is done in the next subsection on dynamic connectedness

Figure 16: Posterior of the indicator influence indices



Posterior distribution of the rolling influence index for the indicator dimension.

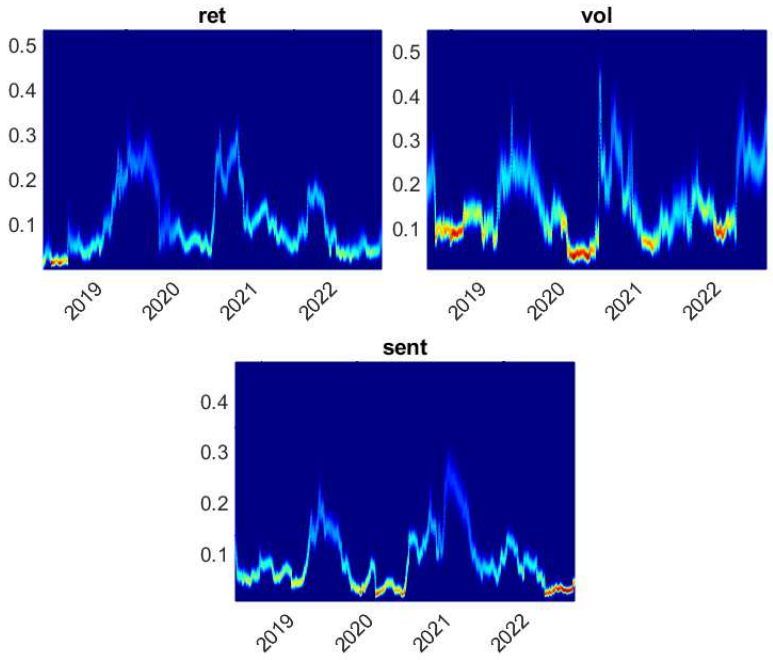
estimation.

Dynamic connectedness

In this subsection, we investigate connectedness measures derived via a dynamic estimation of the proposed MAR model. In particular, we set a rolling window $w = 150$ days (with a total of 1583 estimation stages), where each one has a fixed lag length of $P = 2$, and a forecast horizon for the dynamic analysis of 20-step-ahead.

We first illustrate in Figure 15 the posterior distribution of the dynamic

Figure 17: Posterior of the indicator dependence indices

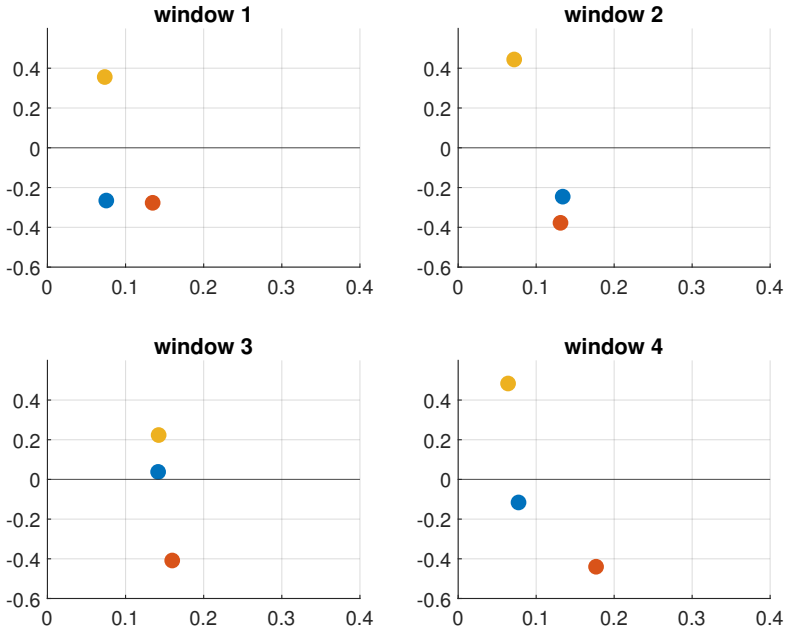


Posterior distribution of the rolling dependence index for the indicator dimension.

total spillover index for the two dimensions: indicator and crypto dimension.

It is evident that the indicator dimension exhibit comparably lower magnitudes of total spillovers compared to the crypto dimension, though there are some phases of relatively increased/decreased interconnectedness. Interestingly, there is no apparent synchronicity between the levels of the two spillover indices: an increased spillover in the indicator level is not associated with neither an increased nor a decreased level of spillover in the crypto dimension. However, during these turbulence phases, both posterior estimates show a high degree of variability.

Figure 18: Scatter of the influence-dependence indices



Indicator influence-dependence scatter plot for the four windows. Colours represent different indicators: blue (return), orange (volatility), yellow (sentiment).

Indicator analysis

The indicator analysis studies interconnectedness across different indicators belonging to the same cryptocurrency. We show in Figure 14 the dynamic evolution of the posterior median of the GFEVD for each equation. There seems to be a relative independence across the three indicators, although there are some periods in which spillovers rise. Returns and, particularly, sentiment transmit information to the system in such cases, whereas volatility has a limited informational outflow.

The latter considerations are corroborated by the influence and depen-

dence indices depicted in Figures 16 and 17, respectively. Sentiment appears a relevant net shock transmitter, volatility as a net shock receiver, with returns exhibiting similar magnitude of from and to spillovers. Dependence indices are comparable in their magnitudes, with the one related to volatility showing relatively higher values.

To gain further insights on the dynamic relationships across indicators, we represent each indicator in the influence-dependence space in Figure 18. We select four equally sized windows of 13 months covering our sample period, i.e. 1 June 2018 - 30 June 2019 (window 1), 1 July 2019 - 30 July 2020 (window 2), 1 August 2020 - 30 August 2021 (window 3), 1 September 2021 - 30 September 2022 (window 4). In this way we can determine the extent to which each indicator can be viewed as dependent or independent from the dynamics of the others (from spillover higher or lower than the within spillover), as well as shock transmitter or receiver (to spillover higher or lower than from spillover).

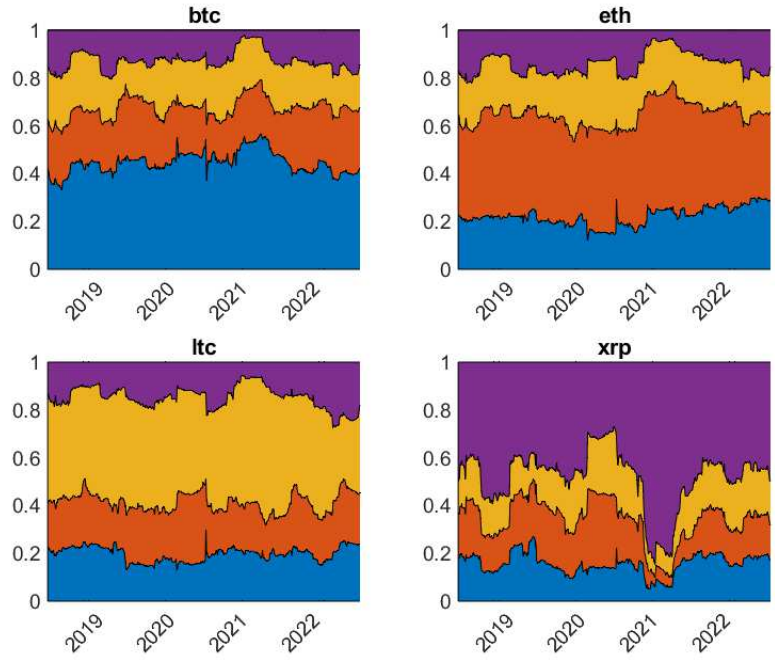
Sentiment is systematically located in the upper-left section of the space, thus being the most influencing variable, with a strong within spillover that makes it resilient to shocks transmitted from others. Similarly, returns exhibit a stronger within spillover than the spillover received from others, though the spillovers transmitted to others are of lower magnitude than the ones received from others - except for the third phase.

Volatility is instead placed in the lower part of the space, with some phases showing a relatively higher dependence index than those of others, suggesting a strong dependence combined with a low influence towards the others

Crypto analysis

The crypto analysis investigates interconnectedness across the same indicators belonging to different cryptocurrencies. We show in Figure 19 the

Figure 19: Posterior distribution of the dynamic cryptos GFEVD

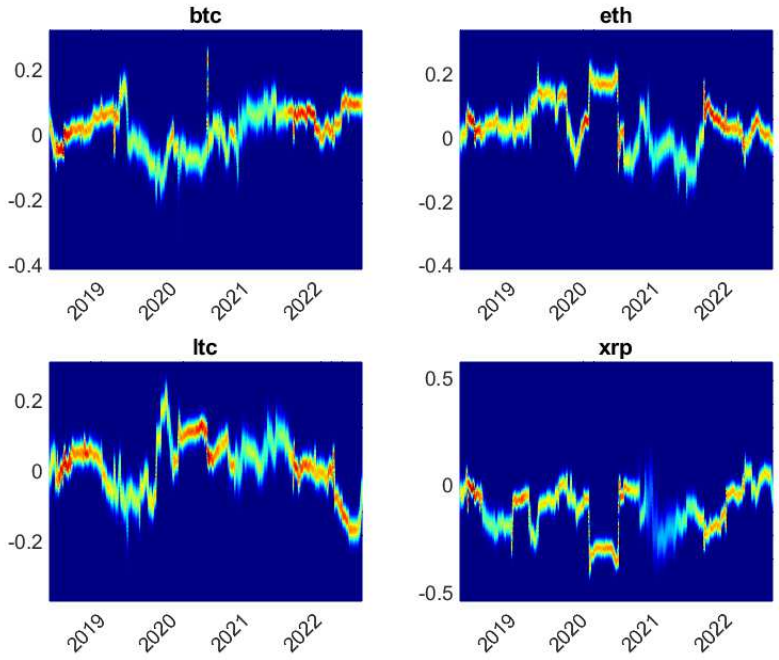


Posterior median of the rolling GFEVD for the crypto dimension.

dynamic evolution of the posterior median of the GFEVD related to the crypto dimension. As also confirmed by the total spillovers, the crypto dimension shows a much higher degree of interconnectedness compared to the variable dimension. Although own variance shares concur to a larger extent to spillovers, cross-variance shares are of considerable magnitude, relatively stable, and generally weigh for more than half of each single crypto GFEVD. An exception is Ripple, whose portion of own variance is higher than those of the other cryptocurrencies, and particularly in the aftermath of 2021.

The posterior distribution of the dynamic influence and dependence indices for each cryptocurrency are represented in Figures 20 and 21, respectively. On the one hand, the posterior median influences appear more volatile

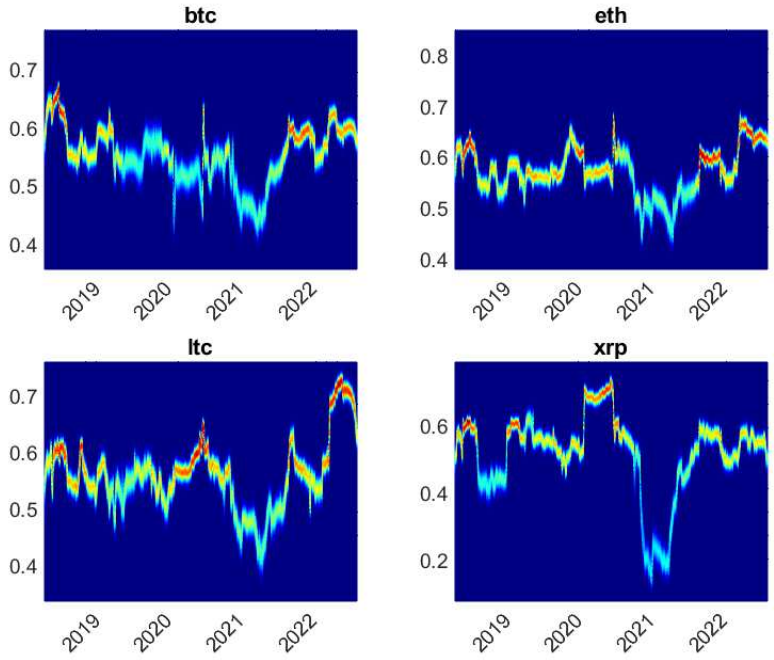
Figure 20: Posterior of the crypto influence indices



Posterior distribution of the rolling influence index for the indicator dimension.

than their indicator counterparts, as well as they cross the zero horizontal line more frequently. This suggests an as such more frequent change in shock transmission leadership across cryptocurrencies, without the possibility to rank them systematically in terms of dominance. The only exception is given by Ripple, which consistently manifests as an overall receiver of shock impacts. On the other hand, the posterior median dependence indices show a relatively steady dynamics, as well as similarity of magnitudes. Despite that, during the period following 2021, all dependence indices drop significantly, due to an increased within spillover dynamics. This affects Ripple in a more noticeable way, with the dependence index dropping by more than two thirds with respect to its initial sample values.

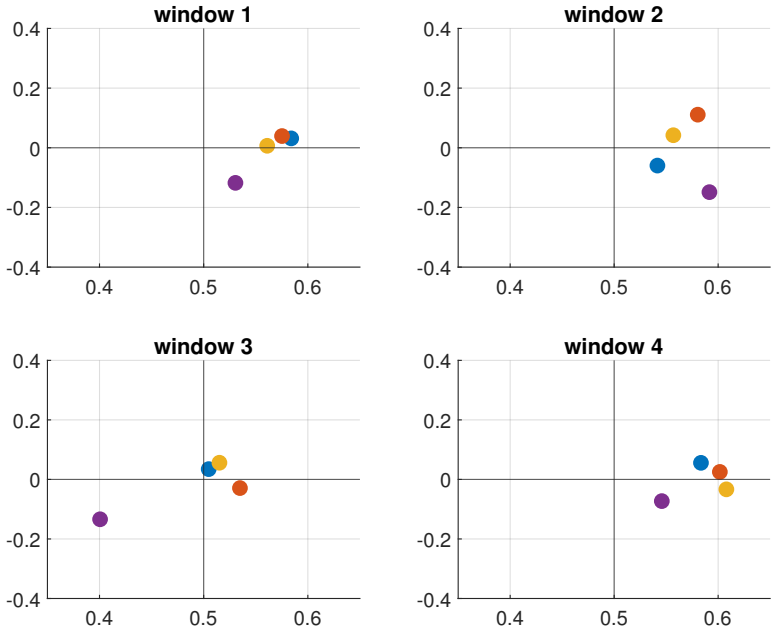
Figure 21: Posterior of the crypto dependence indices



Posterior distribution of the rolling dependence index for the indicator dimension.

As for the indicator case, we represent each cryptocurrency in the influence-dependence space in Figure 22. The joint representation of influence-dependence confirms the non-existence of a clear ranking of cryptocurrencies in terms of their shock transmission capability. Particularly, the leadership seems to be shared by Bitcoin, Ethereum and Litecoin, but this comes without a marked dominance. However, evidence suggests that Ripple does not appear as a net shock transmitter and, even more, it is generally less dependent on the others, except during the second phase. Further, all cryptocurrencies become more independent during the third phase, confirming the previous finding of a more asynchronous dynamics of the indicators within each cryptocurrency.

Figure 22: Scatter of the influence-dependence cryptos



Indicator influence-dependence scatter plot for the four windows. Colours represent different indicators: blue (return), orange (volatility), yellow (sentiment).

4.5 Forecasting Experiment

We now illustrate the forecasting performance of our model in an out-of-sample rolling exercise against some competitor models, which we present below.

Along with the MAR, we consider four models which arise by imposing a priori restrictions on the VAR. Indeed, standard time series econometric theory imposes to work with the vectorized version of our dependent variable $\mathbf{y}_t = \text{vec}(\mathbf{Y}_t)$. In a fully parametrized VAR framework we have, for each lag order, a high-dimensional coefficient matrix $\mathbf{\Phi}$ embedding G^2N^2 parameters.

Nevertheless, when the number of observations is limited, as in our dynamic connectedness analysis, the standard ML estimator is inefficient and will work poorly in a forecasting exercise. Thus, the unrestricted VAR is likely to be an inadequate informative benchmark. Here we do not employ data-driven standard dimensionality reduction techniques mainly because they imply hyperparameter tuning which makes the estimation practically infeasible, provided the large number of rolling window estimations. We stress that the MAR approach is a natural consequence of treating variables as matrices and thus is hyperparameter-free.

We now explain how the different restrictions work. Without loss of generality, we restrict the attention to the case $P = 1$. Higher-order lags are just a generalization of the first-order case. Let us define $\mathbf{\Phi}_{ij}$ to be the $G \times G$ block of $\mathbf{\Phi}$ providing the response of cryptocurrency i to the first lag of cryptocurrency j. For matrix $\mathbf{\Phi}_{ij}$, the coefficient ϕ_{ij}^{lm} gives the response of indicator l of cryptocurrency i to the first lag of indicator m of cryptocurrency j. From the least to the most restricted case, we have four scenarios:

1. VAR: no restrictions are imposed. $\phi_{ij}^{lm} \neq 0 \ \forall \ i, j, l, m$. The model has a total of G^2N^2 parameters.

2. Indicator block VAR (IB-VAR): there are no feedback effects among indicators, but all the cryptocurrencies are allowed to affect each other. $\phi_{ij}^{lm} \neq 0 \ \forall \ i, j$ but only if $l = m$. The model has a total of GN^2 parameters. This corresponds to performing N VAR estimations, one for each cryptocurrency's set of indicators.

3. Crypto block VAR (CB-VAR): there are no feedback effects between different cryptocurrencies, but all the indicators affect each other. $\phi_{ij}^{lm} \neq 0 \ \forall \ l, m$ but only if $i = j$. The model has a total of NG^2 parameters. This corresponds to performing a G VAR estimations, one for each indicator belonging to a set of cryptocurrencies.

4. AR: there are only univariate autoregressive dynamics. $\phi_{ij}^{lm} \neq 0$ only if

$i = j$ and $l = m$. The model has just GN parameters.

We estimate the models with same setting of the dynamic analysis. We then forecast each variable out-of-sample, with step ahead forecasts $H = 1, ..., 4$. Forecast accuracy is then measured in terms of Mean Squared Forecast Error (MSFE). In particular, let $\mathcal{M} = 1, ..., 4$ be the set of models considered, with $\mathcal{M} = 0$ being the MAR model as a benchmark. For each ending point $T = 150, ..., 1733$ of the rolling windows, the H-step-ahead MSFE for variable $d = 1, ..., D$ of the $\mathcal{M}$-th model is defined as:

$$MSFE^{\mathcal{M}}_{d,T,H} = \frac{1}{H}\sum_{h=1}^{H}(\hat{y}^{\mathcal{M}}_{d,T+h} - y_{d,T+h})^2 \tag{4.22}$$

where $\hat{y}^{\mathcal{M}}_{d,T+h}$ denotes the h-step forecast of indicator d of model $\mathcal{M}$, whereas $y_{d,T+h}$ is the corresponding observed value. We report results for MSFE relative to the benchmark, i.e. the relative MSFE (RMSFE), that is:

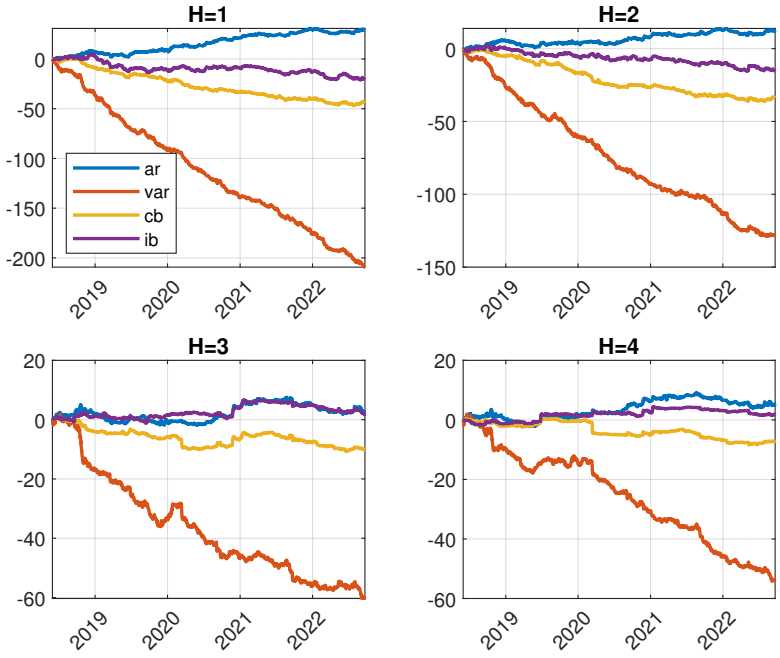
$$RMSFE^{\mathcal{M}}_{d,T,H} = \frac{MSFE^{0}_{d,T,H}}{MSFE^{\mathcal{M}}_{d,T,H}} \tag{4.23}$$

A value greater than 0 means that model $\mathcal{M}$ exhibit a better forecasting performance compared to the MAR, and viceversa.

The results of the forecasting exercise can be graphically seen in Figures 23, 24, and 25, where the cumulative logarithms of the RMSFE are depicted, for each variable, for all the forecast horizons. To ease the interpretation, results are presented as averages for the four cryptocurrencies considered. We refer the reader to Figure 35 in Appendix for a comparative representation of the Bayes-Schwarz information criteria for the MAR model and the analyzed competing alternatives. Additionally, results of the forecasting experiment with alternative rolling window settings ($w = 100, 150, 200$) and lag order ($P = 1, 2$) are reported in Figures 36-38.

Overall, we observe that the VAR model has a poor forecasting power with respect to the extant competing alternatives. This can be reconducted

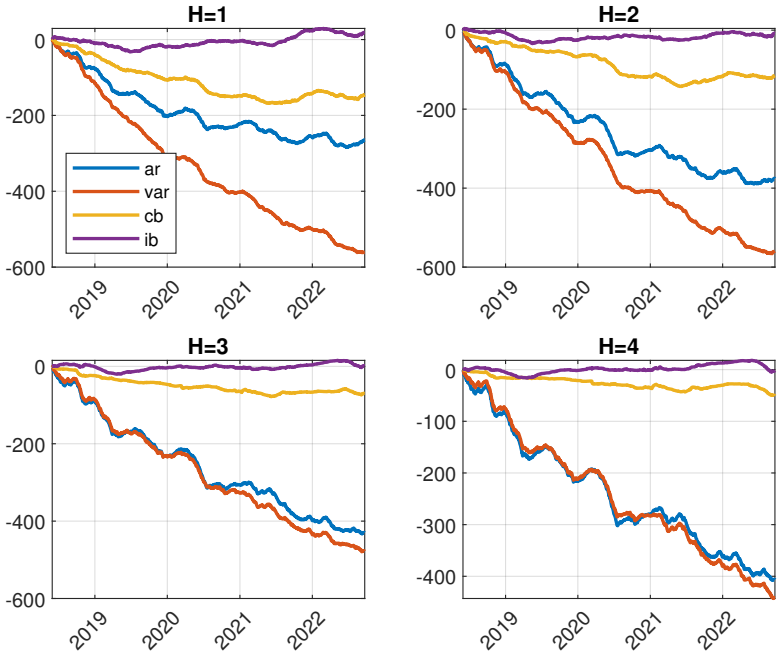
Figure 23: Cumulative logarithm of the RMSFE for the return indicator.



to the current high-dimensional setting for which we are evaluating forecasts. Another interesting common finding is that the IB-VAR generally outperforms the CB-VAR. This result is the offspring of the fact that, as far as forecasting performance is concerned, it is better to model in a VAR fashion the same indicator belonging to different cryptocurrencies, rather than the set of different indicators pertaining to the same cryptocurrency. The latter conclusion supports once more the relatively higher interrelated dynamics of the crypto dimension rather than that of the indicator dimension.

When considering forecasting performances in detail, we observe a different behaviour of the predictive accuracy of the considered models depending on the indicator to be forecasted. As far as price returns, the MAR model generally outperforms the VAR, CB-VAR and IB-VAR, while the latter yields sometime comparable values of the RMSFE. The AR, instead,

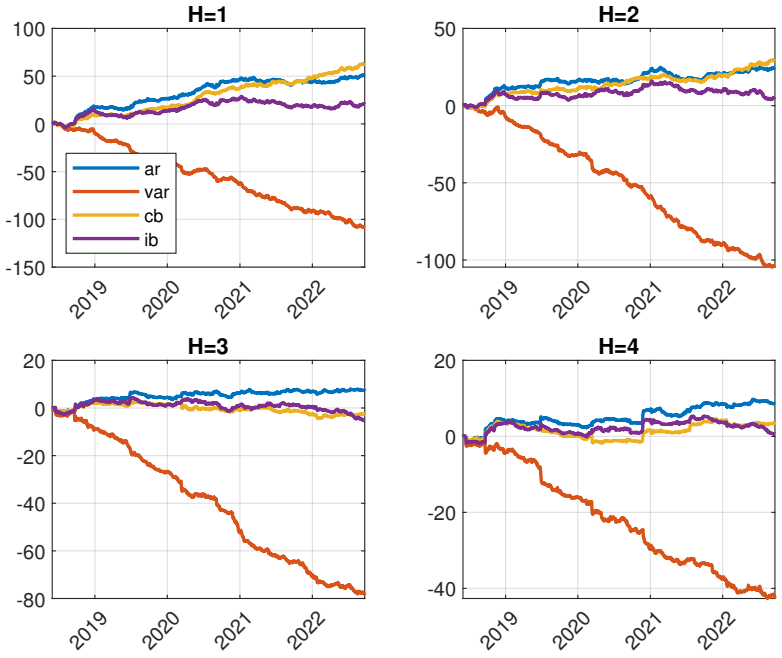
Figure 24: Cumulative logarithm of the RMSFE for the volatility indicator.



tends to perform better than all other alternative models: this is generally true, though in some cases it has similar forecasting accuracy with respect to the MAR model. As far as volatility is concerned, the MAR model generally outperforms all competing models. While the IB-VAR is close in terms of forecasting accuracy, the CB-VAR and, especially, the AR and VAR models perform worse. Sentiment is instead the variable for which the MAR model yields poorer forecasting performances, if compared to the others - except for the VAR model. In this case, the AR model overperforms the competitors for forecast horizons 3 and 4, whereas it yields quite comparable forecast accuracy with the two block models, particularly with the IB-VAR.

We further perform the Diebold and Mariano (2002) test for predictive accuracy to compare the forecasts given by the MAR against the AR and VAR models, whose results are contained in Tables 6 and 7 in Appendix.

Figure 25: Cumulative logarithm of the RMSFE for the sentiment indicator.



The test shows that, overall, the VAR and the MAR model generally exhibit different forecasting accuracy, which confirms the superior performance of the MAR model in terms of forecasting. When it comes to the comparison of MAR and AR, the test often does not reject the null hypothesis of a zero expected loss differential and, particularly, when returns are forecasted.

4.6 Conclusion

We have proposed a ML and Bayesian estimation of autoregressive models for matrix-valued financial time series. The model, compared to a standard vector autoregressive model, preserves the original matrix structure of financial data and admits corresponding enhanced interpretations due to its bilinear form. We have outlined a procedure to derive generalized multidimensional

connectedness measures, and provided a forecasting experiment which highlights the comparable forecasting performances of the model relative to AR and VAR models.

Our approach and its results might be useful for risk management and regulatory policy making: while fund managers aim mostly at predicting returns, regulatory institutions are periodically asked to forecast volatility, whilst sentiment being a potentially key predictor.

As supported by the theoretical literature, our findings reveal only a weak evidence of directional spillover from sentiment indicators to returns and volatility. Therefore, there is weak evidence towards the hypothesis that noise traders impact returns or volatility.

There are many possible extensions to the outlined approach. In the future, time-varying parameter autoregressive models for matrix-valued time series could be considered. Further, dynamic factor matrix autoregression can be of interest for multidimensional market models.

5 Concluding Remarks

The motivation of this thesis is grounded on the growing attention of statistical and econometric methods for high and multidimensional frameworks. In conclusion, this thesis makes a significant contribution to the literature of high and multi-dimensional models in the fields of Economics and Finance. By proposing novel Bayesian estimation techniques and Graph Theory approaches, we aim to address the inferential and interpretational issues that arise when modeling complex systems.

In Chapter 2 we studied shock propagation in longitudinal time series data using statistical and econometric techniques and network theory. We build on a dynamic spillover network framework derived from a Global VAR model, and apply network analysis to study the network topology of spillover indices across European countries, macroeconomic variables, and sentiment data. The usefulness of the method is demonstrated by analyzing the network structure of shock propagation and showing how different crises have diverse impacts on the real economy, economic sentiment, and the entire countries' macroeconomic network structure.

In Chapter 3 we proposed a Bayesian method for variable selection in high-dimensional MAR models. The method consists of two computational procedures: an MCMC algorithm and an EMVS estimation procedure, which allow for fast dynamic posterior exploration. The efficiency of the proposed estimators was numerically investigated, and both are found to perform well compared to alternative models. The method was applied to a panel of nine countries' economic indicators. We find that US and Canada have the largest pairwise contributions to each other's economy, while both the US and China are prone to transmitting large portions of GFEVD to other countries, with China generally being more resilient to shocks in other countries.

In Chapter 4 we proposed a method for estimating autoregressive models for financial time series using both maximum likelihood and Bayesian tech-

niques. The method preserves the original matrix structure of the financial data, allowing for enhanced interpretations due to its bilinear form, and also includes a procedure to derive generalized multidimensional connectedness measures. The forecasting experiment showed that the proposed model had comparable forecasting performance relative to AR and VAR models. The results of the study have potential applications for risk management and regulatory policy-making. The findings reveal only weak evidence of directional spillover from sentiment indicators to returns and volatility, suggesting limited impact of noise traders on returns or volatility.

Extensions and Further Research

While the methods described in Chapter 2 are an important step towards combining econometrics and network theory, there is still a lot of room for further development in both methodology and practical applications in these fields. Currently, network theory is not often applied to econometric models, but it has the potential to greatly enhance our understanding of time series models and to facilitate the creation of interdisciplinary research fields.

The most promising evolution of the Variable Selection procedure for MAR presented in Chapter 3 is probably the inclusion of time variation in the coefficients. This extension would enable the modeling of more sophisticated and practical scenarios, particularly in the financial and economic domains, which would result in more precise and practical outcomes.

From the methodological point of view, the possible extensions concerning the MAR employed in Chapter 4, basically resemble the one suited for the Variable Selection MAR. Nevertheless, an intriguing empirical application would be to consider an enlarged Financial dataset, where different crypto prices are observed through different exchanges. This would also lay the ground for the development of cointegrated methods for matrix and tensor variate data. Different exchange prices obey to the one price law, which

annilihates the possibility of arbitrage by buying and selling through different exchanges. Now suppose that data are organized for each time point as a matrix with rows representing exchanges and colums cryptos. In this case, the left hand autoregressive matrices could be reformulated as mimic a VECM, whereas at the same time the right hand matrices are free from any restriction.

References

[1] Ahelegbey, D., M. Billio, and R. Casarin (2016). Bayesian graphical models for structural vector autoregressive processes. Journal of Applied Econometrics 31(2), 357–386.

[2] Ahelegbey, D. F., P. Cerchiello, and R. Scaramozzino (2022). Network based evidence of the financial impact of covid-19 pandemic. International Review of Financial Analysis 81, 102101.

[3] Ahelegbey, D. F., E. S. Fianu, and L. Grossi (2022). Modeling risk contagion in the italian zonal electricity market. European Journal of Operational Research 298(2), 656–679.

[4] Ahelegbey, D. F. and P. Giudici (2022). Netvix—a network volatility index of financial markets. Physica A: Statistical Mechanics and its Applications 594, 127017.

[5] Aldasoro, I. and I. Alves (2018). Multiplex interbank networks and systemic importance: An application to european data. Journal of Financial Stability 35, 17–37.

[6] Alizadeh, S., M. W. Brandt, and F. X. Diebold (2002). Range-based estimation of stochastic volatility models. The Journal of Finance 57(3), 1047–1091.

[7] Antweiler, W. and M. Z. Frank (2004). Is all that talk just noise? the information content of internet stock message boards. The Journal of finance 59(3), 1259–1294.

[8] Anyfantaki, S., S. Arvanitis, and N. Topaloglou (2021). Diversification benefits in the cryptocurrency market under mild explosivity. European Journal of Operational Research 295(1), 378–393.

[9] Aouadi, A., M. Arouri, and F. Teulon (2013). Investor attention and stock market activity: Evidence from france. Economic Modelling 35, 674–681.

[10] Asness, C. S., T. J. Moskowitz, and L. H. Pedersen (2013). Value and momentum everywhere. The Journal of Finance 68(3), 929–985.

[11] Atsalakis, G. S., I. G. Atsalaki, F. Pasiouras, and C. Zopounidis (2019). Bitcoin price forecasting with neuro-fuzzy techniques. European Journal of Operational Research 276(2), 770–780.

[12] Bańbura, M., D. Giannone, and L. Reichlin (2010). Large bayesian vector auto regressions. Journal of Applied Econometrics 25(1), 71–92.

[13] Bai, J. and S. Ng (2002). Determining the number of factors in approximate factor models. Econometrica 70(1), 191–221.

[14] Barber, B. M. and T. Odean (2008). All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors. The review of financial studies 21(2), 785–818.

[15] Barber, B. M., T. Odean, and N. Zhu (2009). Systematic noise. Journal of Financial Markets 12(4), 547–569.

[16] Bardoscia, M., S. Battiston, F. Caccioli, and G. Caldarelli (2017). Pathways towards instability in financial networks. Nature Communications 8(1), 1–7.

[17] Bartolucci, F., F. Pennoni, and A. Mira (2021). A multivariate statistical approach to predict covid-19 count data with epidemiological interpretation and uncertainty quantification. Statistics in Medicine 40(24), 5351–5372.

[18] Battiston, S. et al. (2019). The importance of being forward-looking: managing financial stability in the face of climate risk. Financial Stability Review (23), 39–48.

[19] Battiston, S., G. Caldarelli, R. M. May, T. Roukny, and J. E. Stiglitz (2016). The price of complexity in financial networks. Proceedings of the National Academy of Sciences 113(36), 10031–10036.

[20] Battiston, S., A. Mandel, I. Monasterolo, F. Schütze, and G. Visentin (2017). A climate stress-test of the financial system. Nature Climate Change 7(4), 283–288.

[21] Battiston, S., I. Monasterolo, K. Riahi, and B. J. van Ruijven (2021). Accounting for finance is key for climate mitigation pathways. Science 372(6545), 918–920.

[22] Behrendt, S. and A. Schmidt (2018). The twitter myth revisited: Intraday investor sentiment, twitter activity and individual-level stock return volatility. Journal of Banking & Finance 96, 355–367.

[23] Billio, M., R. Casarin, M. Iacopini, and S. Kaufmann (2022). Bayesian dynamic tensor regression. Journal of Business & Economic Statistics, 1–11.

[24] Billio, M., M. Getmansky, A. W. Lo, and L. Pelizzon (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. Journal of financial economics 104(3), 535–559.

[25] Bitetto, A., P. Cerchiello, and C. Mertzanis (2021). A data-driven approach to measuring epidemiological susceptibility risk around the world. Scientific Reports 11.

[26] Bleher, J. and T. Dimpfl (2019). Today i got a million, tomorrow, i don't know: on the predictability of cryptocurrencies by means of google search volume. International Review of Financial Analysis 63, 147–159.

[27] Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre (2008). Fast unfolding of communities in large networks. Journal of statistical mechanics: theory and experiment 2008(10), P10008.

[28] Bonacich, P. (2007). Some unique properties of eigenvector centrality. Social networks 29(4), 555–564.

[29] Brin, S. and L. Page (1998). The anatomy of a large-scale hypertextual web search engine. Computer networks and ISDN systems 30(1-7), 107–117.

[30] Brown, G. W. and M. T. Cliff (2004). Investor sentiment and the near-term stock market. Journal of empirical finance 11(1), 1–27.

[31] Burriel, P. and A. Galesi (2018). Uncovering the heterogeneous effects of ecb unconventional monetary policies across euro area countries. European Economic Review 101, 210–229.

[32] Canova, F. and M. Ciccarelli (2009). Estimating multicountry var models. International Economic Review 50(3), 929–959.

[33] Canova, F. and M. Ciccarelli (2013). Panel vector autoregressive models: A survey. Advances in Econometrics 31.

[34] Caporin, M. and F. Poli (2022). News and intraday jumps: Evidence from regularization and class imbalance. The North American Journal of Economics and Finance 62, 101743.

[35] Catania, L., S. Grassi, and F. Ravazzolo (2019). Forecasting cryptocurrencies under model and parameter instability. International Journal of Forecasting 35(2), 485–501.

[36] Celani, A. and P. Pagnottoni (2022). Matrix autoregressive models: Generalization and bayesian estimation. Available at SSRN.

[37] Celani, A. and P. Pagnottoni (2023). The multidimensional relationships between sentiment, returns and volatility. Available at SSRN.

[38] Cheah, E.-T. and J. Fry (2015). Speculative bubbles in bitcoin markets? an empirical investigation into the fundamental value of bitcoin. Economics letters 130, 32–36.

[39] Chen, E. Y. and J. Fan (2021). Statistical inference for high-dimensional matrix-variate factor models. Journal of the American Statistical Association, 1–18.

[40] Chen, R., W. Bao, and C. Jin (2021). Investor sentiment and predictability for volatility on energy futures markets: Evidence from china. International Review of Economics & Finance 75, 112–129.

[41] Chen, R., X. H., and D. Yang (2021). Autoregressive models for matrix-valued time series. Journal of Econometrics 222(1, Part B), 539–560.

[42] Chen, R., D. Yang, and C.-H. Zhang (2022). Factor models for high-dimensional tensor time series. Journal of the American Statistical Association 117(537), 94–116.

[43] Chen, W., H. Xu, L. Jia, and Y. Gao (2021). Machine learning model for bitcoin exchange rate prediction using economic and technology determinants. International Journal of Forecasting 37(1), 28–43.

[44] Chudik, A. and M. H. Pesaran (2013). Econometric analysis of high dimensional vars featuring a dominant unit. Econometric Reviews 32(5-6), 592–649.

[45] Chudik, A. and M. H. Pesaran (2016). Theory and practice of gvar modelling. Journal of Economic Surveys 30(1), 165–197.

[46] Cichocki, A. (2018). Fundamental tensor operations for large-scale data analysis using tensor network formats. Multidimensional Systems and Signal Processing 29, 921–960.

[47] Cuaresma, J. C., M. Feldkircher, and F. Huber (2016). Forecasting with global vector autoregressive models: a bayesian approach. Journal of Applied Econometrics 31(7), 1371–1391.

[48] Da, Z., J. Engelberg, and P. Gao (2011). In search of attention. The journal of finance 66(5), 1461–1499.

[49] Da, Z., J. Engelberg, and P. Gao (2015). The sum of all fears investor sentiment and asset prices. The Review of Financial Studies 28(1), 1–32.

[50] Dafermos, Y., M. Nikolaidi, and G. Galanis (2018). Climate change, financial stability and monetary policy. Ecological Economics 152, 219–234.

[51] De Long, J. B., A. Shleifer, L. H. Summers, and R. J. Waldmann (1990). Noise trader risk in financial markets. Journal of political Economy 98(4), 703–738.

[52] Dees, S., F. d. Mauro, M. H. Pesaran, and L. V. Smith (2007). Exploring the international linkages of the euro area: a global var analysis. Journal of applied econometrics 22(1), 1–38.

[53] Delis, M. D., C. S. Savva, and P. Theodossiou (2021). The impact of the coronavirus crisis on the market price of risk. Journal of Financial Stability 53, 100840.

[54] Diebold, F. X. and R. S. Mariano (2002). Comparing predictive accuracy. Journal of Business & economic statistics 20(1), 134–144.

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

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

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

[58] Diebold, F. X. and K. Yilmaz (2015). Financial and macroeconomic connectedness: A network approach to measurement and monitoring. Oxford University Press, USA.

[59] Dimpfl, T. and S. Jank (2016). Can internet search queries help to predict stock market volatility? European financial management 22(2), 171–192.

[60] Dimpfl, T. and F. J. Peter (2021). Nothing but noise? price discovery across cryptocurrency exchanges. Journal of Financial Markets 54, 100584.

[61] El Montasser, G., L. Charfeddine, and A. Benhamed (2022). Covid-19, cryptocurrencies bubbles and digital market efficiency: sensitivity and similarity analysis. Finance Research Letters 46, 102362.

[62] Elhorst, J. P., M. Gross, and E. Tereanu (2021). Cross-sectional dependence and spillovers in space and time: where spatial econometrics and global var models meet. Journal of Economic Surveys 35(1), 192–226.

[63] Fang, L. and J. Peress (2009). Media coverage and the cross-section of stock returns. The Journal of Finance 64(5), 2023–2052.

[64] Fisher, K. L. and M. Statman (2000). Investor sentiment and stock returns. Financial Analysts Journal 56(2), 16–23.

[65] Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2005). The generalized dynamic factor model. Journal of the American Statistical Association 100(471), 830–840.

[66] Fosten, J. and R. Greenaway-McGrevy (2022). Panel data nowcasting. Econometric Reviews 41(7), 675–696.

[67] Gao, Z. and R. S. Tsay (2021). A two-way transformed factor model for matrix-variate time series. Econometrics and Statistics.

[68] Garman, M. B. and M. J. Klass (1980). On the estimation of security price volatilities from historical data. Journal of business, 67–78.

[69] Gefang, D. (2014). Bayesian doubly adaptive elastic-net lasso for var shrinkage. International Journal of Forecasting 30(1), 1–11.

[70] George, E. I. and R. E. McCulloch (1993). Variable selection via gibbs sampling. Journal of the American Statistical Association 88(423), 881–889.

[71] George, E. I. and R. E. McCulloch (1997). Approaches for bayesian variable selection. Statistica sinica, 339–373.

[72] George, E. I., D. Sun, and S. Ni (2008). Bayesian stochastic search for var model restrictions. Journal of Econometrics 142(1), 553–580.

[73] Geyer, C. J. (2010). Computation for the introduction to mcmc chapter of handbook of markov chain monte carlo.

[74] Gong, L. and J. M. Flegal (2016). A practical sequential stopping rule for high-dimensional markov chain monte carlo. Journal of Computational and Graphical Statistics 25(3), 684–700.

[75] Gradojevic, N., D. Kukolj, R. Adcock, and V. Djakovic (2021). Forecasting bitcoin with technical analysis: A not-so-random forest? International Journal of Forecasting.

[76] Greenwood-Nimmo, M., V. H. Nguyen, and Y. Shin (2021). Measuring the connectedness of the global economy. International Journal of Forecasting 37(2), 899–919.

[77] Gross, M. and C. Kok (2013). Measuring contagion potential among sovereigns and banks using a mixed-cross-section gvar.

[78] Grullon, G., G. Kanatas, and J. P. Weston (2004). Advertising, breadth of ownership, and liquidity. The Review of Financial Studies 17(2), 439–461.

[79] Gu, C. and A. Kurov (2020). Informational role of social media: Evidence from twitter sentiment. Journal of Banking & Finance 121, 105969.

[80] Gupta, A. and N. D.K. (1999). Matrix variate distributions. Chapman & Hall/CRC.

[81] Hamid, A. and M. Heiden (2015). Forecasting volatility with empirical similarity and google trends. Journal of Economic Behavior & Organization 117, 62–81.

[82] Harbo, I., S. Johansen, B. Nielsen, and A. Rahbek (1998). Asymptotic inference on cointegrating rank in partial systems. Journal of Business & Economic Statistics 16(4), 388–399.

[83] Harris, J., J. L. Hirst, and M. Mossinghoff (2008). Combinatorics and Graph Theory. Springer New York.

[84] Harris, M. and A. Raviv (1993). Differences of opinion make a horse race. The Review of Financial Studies 6(3), 473–506.

[85] Hayes, A. S. (2019). Bitcoin price and its marginal cost of production: support for a fundamental value. Applied economics letters 26(7), 554–560.

[86] Ho, K.-Y., Y. Shi, and Z. Zhang (2013). How does news sentiment impact asset volatility? evidence from long memory and regime-switching approaches. The North American Journal of Economics and Finance 26, 436–456.

[87] Hoff, P. D. (2011). Separable covariance arrays via the Tucker product, with applications to multivariate relational data. Bayesian Analysis 6(2), 179 – 196.

[88] Hoff, P. D. (2015). Multilinear tensor regression for longitudinal relational data. The Annals of Applied Statistics 9(3), 1169–1193.

[89] Iwanicz-Drozdowska, M., K. Rogowicz, Ł. Kurowski, and P. Smaga (2021). Two decades of contagion effect on stock markets: Which events are more contagious? Journal of Financial Stability 55, 100907.

[90] Jones, G. L. and Q. Qin (2022). Markov chain monte carlo in practice. Annual Review of Statistics and Its Application 9(1), 557–578.

[91] Joseph, K., M. B. Wintoki, and Z. Zhang (2011). Forecasting abnormal stock returns and trading volume using investor sentiment: Evidence from online search. International Journal of Forecasting 27(4), 1116–1127.

[92] Kapar, B. and J. Olmo (2021). Analysis of bitcoin prices using market and sentiment variables. The World Economy 44(1), 45–63.

[93] Katz, L. (1953). A new status index derived from sociometric analysis. Psychometrika 18(1), 39–43.

[94] Kleinberg, J. M. (1998). Authoritative sources in a hyperlinked environment. In Proceedings of the ninth annual ACM-SIAM symposium on Discrete algorithms, pp. 668–677.

[95] Kleinberg, J. M. (1999, sep). Authoritative sources in a hyperlinked environment. J. ACM 46(5), 604–632.

[96] Koch, S. and T. Dimpfl (2023). Attention and retail investor herding in cryptocurrency markets. Finance Research Letters 51, 103474.

[97] Kock, A. and L. Callot (2015). Oracle inequalities for high dimensional vector autoregressions. Journal of Econometrics 186(2), 325–344.

[98] Kolda, T. G. and B. W. Bader (2009, September). Tensor decompositions and applications. SIAM Review 51(3), 455–500.

[99] Konstantakis, K. N., P. G. Michaelides, E. G. Tsionas, and C. Minou (2015). System estimation of gvar with two dominants and network theory: Evidence for brics. Economic Modelling 51, 604–616.

[100] Koop, G. and D. Korobilis (2013). Large time-varying parameter vars. Journal of Econometrics 177(2), 185–198. Dynamic Econometric Modeling and Forecasting.

[101] Koop, G. and D. Korobilis (2016). Model uncertainty in panel vector autoregressive models. European Economic Review 81, 115–131. Model Uncertainty in Economics.

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

[103] Korobilis, D. (2021). High-dimensional macroeconomic forecasting using message passing algorithms. Journal of Business & Economic Statistics 39(2), 493–504.

[104] Kose, M. A., C. Otrok, and C. H. Whiteman (2003). International business cycles: World, region, and country-specific factors. american economic review 93(4), 1216–1239.

[105] Kose, M. A., C. Otrok, and C. H. Whiteman (2008). Understanding the evolution of world business cycles. Journal of international Economics 75(1), 110–130.

[106] Koutmos, D. and J. E. Payne (2021). Intertemporal asset pricing with bitcoin. Review of Quantitative Finance and Accounting 56(2), 619–645.

[107] Kraaijeveld, O. and J. De Smedt (2020). The predictive power of public twitter sentiment for forecasting cryptocurrency prices. Journal of International Financial Markets, Institutions and Money 65, 101188.

[108] Kristoufek, L. (2020). Bitcoin and its mining on the equilibrium path. Energy Economics 85, 104588.

[109] Lam, C., Q. Yao, and N. Bathia (2011, 10). Estimation of latent factors for high-dimensional time series. Biometrika 98(4), 901–918.

[110] Lanne, M. and H. Nyberg (2016). Generalized forecast error variance decomposition for linear and nonlinear multivariate models. Oxford Bulletin of Economics and Statistics 78(4), 595–603.

[111] Liang, C., L. Tang, Y. Li, and Y. Wei (2020). Which sentiment index is more informative to forecast stock market volatility? evidence from china. International Review of Financial Analysis 71, 101552.

[112] Lin, Z.-Y. (2021). Investor attention and cryptocurrency performance. Finance Research Letters 40, 101702.

[113] Litterman, R. B. (1986). Forecasting with bayesian vector autoregressions: Five years of experience. Journal of Business & Economic Statistics 4(1), 25–38.

[114] Liu, Y., B. Qiu, and T. Wang (2021). Debt rollover risk, credit default swap spread and stock returns: Evidence from the covid-19 crisis. Journal of Financial Stability 53, 100855.

[115] Loan, C. F. V. (2000). The ubiquitous kronecker product. Journal of Computational and Applied Mathematics 123(1), 85–100. Numerical Analysis 2000. Vol. III: Linear Algebra.

[116] Lütkepohl, H. (2005). New Introduction to Multiple Time Series Analysis. Springer.

[117] Meng, X.-L. and D. B. Rubin (1993). Maximum likelihood estimation via the ecm algorithm: A general framework. Biometrika 80(2), 267–278.

[118] Micciché, S., G. Bonanno, F. Lillo, and R. N. Mantegna (2003). Degree stability of a minimum spanning tree of price return and volatility. Physica A: Statistical Mechanics and its Applications 324(1-2), 66–73.

[119] Mishkin, F. S. (2011). Over the cliff: From the subprime to the global financial crisis. Journal of Economic Perspectives 25(1), 49–70.

[120] Monasterolo, I., S. Battiston, A. C. Janetos, and Z. Zheng (2017). Vulnerable yet relevant: the two dimensions of climate-related financial disclosure. Climatic Change 145(3), 495–507.

[121] Musciotto, F., L. Marotta, S. Micciché, and R. N. Mantegna (2018). Bootstrap validation of links of a minimum spanning tree. Physica A: Statistical Mechanics and its Applications 512, 1032–1043.

[122] Nakajima, J. (2011). Time-varying parameter var model with stochastic volatility: An overview of methodology and empirical applications. Monetary and and Economic Studies 29, 107–142.

[123] Neal, R. and S. M. Wheatley (1998). Do measures of investor sentiment predict returns? Journal of Financial and Quantitative Analysis 33(4), 523–547.

[124] Neto, D. (2021). Are google searches making the bitcoin market run amok? a tail event analysis. The North American Journal of Economics and Finance 57, 101454.

[125] Ohlson, M., M. Rauf Ahmad, and D. von Rosen (2013). The multilinear normal distribution: Introduction and some basic properties.

Journal of Multivariate Analysis 113, 37–47. Special Issue on Multivariate Distribution Theory in Memory of Samuel Kotz.

[126] Pagnottoni, P., A. Spelta, A. Flori, and F. Pammolli (2022). Climate change and financial stability: Natural disaster impacts on global stock markets. Physica A: Statistical Mechanics and its Applications 599, 127514.

[127] Pagnottoni, P., A. Spelta, N. Pecora, A. Flori, and F. Pammolli (2021). Financial earthquakes: Sars-cov-2 news shock propagation in stock and sovereign bond markets. Physica A: Statistical Mechanics and its Applications 582, 126240.

[128] Park, T. and G. Casella (2008). The bayesian lasso. Journal of the American Statistical Association 103(482), 681–686.

[129] Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004). Modeling regional interdependencies using a global error-correcting macroeconometric model. Journal of Business & Economic Statistics 22(2), 129–162.

[130] Pesaran, M. H. and C. F. Yang (2020). Econometric analysis of production networks with dominant units. Journal of Econometrics 219(2), 507–541.

[131] Philippas, D., H. Rjiba, K. Guesmi, and S. Goutte (2019). Media attention and bitcoin prices. Finance Research Letters 30, 37–43.

[132] Polson, N., J. Scott, B. Clarke, and C. Severinski (2012, January). Shrink Globally, Act Locally: Sparse Bayesian Regularization and Prediction, Volume 9780199694587. Oxford University Press. Publisher Copyright: © Oxford University Press 2011. All rights reserved.

[133] Raftery, A. and S. Lewis (1970, 02). The number of iterations, convergence diagnostics and generic metropolis algorithms. Practical Markov Chain Monte Carlo.

[134] Renault, T. (2017). Intraday online investor sentiment and return patterns in the us stock market. Journal of Banking & Finance 84, 25–40.

[135] Ročková, V. and E. I. George (2014). Emvs: The em approach to bayesian variable selection. Journal of the American Statistical Association 109(506), 828–846.

[136] Ročková, V. and E. I. George (2018). The spike-and-slab lasso. Journal of the American Statistical Association 113(521), 431–444.

[137] Rockova, V. and K. McAlinn (2021). Dynamic variable selection with spike-and-slab process priors. Bayesian Analysis 16(1), 233–269.

[138] Roncoroni, A., S. Battiston, L. O. Escobar-Farfán, and S. Martinez-Jaramillo (2021). Climate risk and financial stability in the network of banks and investment funds. Journal of Financial Stability 54, 100870.

[139] Rothman, A. J., E. Levina, and J. Zhu (2010). Sparse multivariate regression with covariance estimation. Journal of Computational and Graphical Statistics 19(4), 947–962. PMID: 24963268.

[140] Roukny, T., S. Battiston, and J. E. Stiglitz (2018). Interconnectedness as a source of uncertainty in systemic risk. Journal of Financial Stability 35, 93–106.

[141] Schmeling, M. (2007). Institutional and individual sentiment: smart money and noise trader risk? International Journal of Forecasting 23(1), 127–145.

[142] Shen, D., A. Urquhart, and P. Wang (2019). Does twitter predict bitcoin? Economics Letters 174, 118–122.

[143] Shleifer, A. and L. H. Summers (1990). The noise trader approach to finance. Journal of Economic perspectives 4(2), 19–33.

[144] Siganos, A., E. Vagenas-Nanos, and P. Verwijmeren (2017). Divergence of sentiment and stock market trading. Journal of Banking & Finance 78, 130–141.

[145] Sims, C. A. (1980). Macroeconomics and reality. Econometrica 48(1), 1–48.

[146] Sims, C. A. and T. Zha (1998). Bayesian methods for dynamic multivariate models. International Economic Review 39(4), 949–968.

[147] Solt, M. E. and M. Statman (1988). How useful is the sentiment index? Financial Analysts Journal 44(5), 45–55.

[148] Song, S. and P. Bickel (2011, June). Large Vector Auto Regressions. Papers, arXiv.org.

[149] Spelta, A. and P. Pagnottoni (2021). Mobility-based real-time economic monitoring amid the covid-19 pandemic. Scientific Reports 11(1), 1–15.

[150] Stambaugh, R. F., J. Yu, and Y. Yuan (2012). The short of it: Investor sentiment and anomalies. Journal of Financial Economics 104(2), 288–302.

[151] Stolbova, V., I. Monasterolo, and S. Battiston (2018). A financial macro-network approach to climate policy evaluation. Ecological Economics 149, 239–253.

[152] Sun, L., M. Najand, and J. Shen (2016). Stock return predictability and investor sentiment: A high-frequency perspective. Journal of Banking & Finance 73, 147–164.

[153] Tucker, L. R. (1966). Some mathematical notes on three-mode factor analysis. Psychometrika 51, 279–311.

[154] Van Loan, C. F. and N. Pitsianis (1993). Approximation with Kronecker Products, pp. 293–314. Dordrecht: Springer Netherlands.

[155] Van Vliet, B. (2018). An alternative model of metcalfe's law for valuing bitcoin. Economics Letters 165, 70–72.

[156] Vats, D., J. M. Flegal, and G. L. Jones (2019). Multivariate output analysis for Markov chain Monte Carlo. Biometrika 106(2), 321–337.

[157] Vats, D., J. M. Flegal, and G. L. Jones (2021). Monte Carlo Simulation: Are We There Yet?, pp. 1–15. John Wiley & Sons, Ltd.

[158] Wang, D., X. Liu, and R. Chen (2019). Factor models for matrix-valued high-dimensional time series. Journal of Econometrics 208(1), 231–248. Special Issue on Financial Engineering and Risk Management.

[159] Wang, Y.-H., A. Keswani, and S. J. Taylor (2006). The relationships between sentiment, returns and volatility. International Journal of Forecasting 22(1), 109–123.

A Appendix

A.1 Technical Details of Chapter 2

Global VAR forecast error variance decomposition

Given the nature of the variables underlying the Global VAR dynamical system, we follow Greenwood-Nimmo et al. (2021) who propose an aggregation scheme for the GFEVD in order to reduce its dimensionality with a direct interpretation on the countries (variables) FEVD. First, we re-normalise such that $\mathbb{C}_R^{(h)} = K$. After re-normalisation, the (i, j) -th element of $\mathrm{C}_R^{(h)}$ represents the proportion of the total h -steps-ahead FEV of the system accounted for by the spillover effect from variable i to variable i. With this modification we are ensured that we may achieve a percentage interpretation even after aggregating groups of variables in the system.

Suppose we are interested in analyzing the spillover measures developed by Diebold and Yilmaz (2014) but focusing on the countries. It makes sense to aggregate the FEVD according to country blocks, where each ith element is obtained as an aggregation (sum) of the ith country block with its own variables. If, instead, the aim is to carry a variable analysis, the aggregation can be done considering the variable blocks.

Once we have collected $\mathbf{x}_t$ into b groups $\mathbb{C}_R^{(h)}$ can be equivalently expressed as:

$$\underset{(m\times m)}{\mathbb{C}_R^{(h)}} = \begin{bmatrix} \mathbf{B}_{1\leftarrow 1}^{(h)} & \mathbf{B}_{1\leftarrow 2}^{(h)} & \cdots & \mathbf{B}_{1\leftarrow b}^{(h)} \\ \mathbf{B}_{2\leftarrow 1}^{(h)} & \mathbf{B}_{2\leftarrow 2}^{(h)} & \cdots & \mathbf{B}_{2\leftarrow b}^{(h)} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_{b\leftarrow 1}^{(h)} & \mathbf{B}_{b\leftarrow 2}^{(h)} & \cdots & \mathbf{B}_{b\leftarrow b}^{(h)} \end{bmatrix}. \tag{A.1}$$

Let us stress that no information is lost in this process. Consider all the blocks lying on the main diagonal of that matrix (i.e. $\mathbf{B}_{k\leftarrow k}^{(h)}$); they

contain all of the within-group FEV contributions. We can therefore define the within-group FEV contribution for the k th group as follows

$$\mathcal{W}_{k\leftarrow k}^{(h)} = \mathbf{e}'_{m_k}\mathbf{B}_{k\leftarrow k}^{(h)}\mathbf{e}_{m_k}, \tag{A.2}$$

where $\mathbf{e}_{m_k}$ is the usual selection vector. Roughly speaking, the within-group FEV contribution for the k th group is equal to the sum of the elements of the block $\mathbf{B}_{k\leftarrow k'}^{(h)}$. Analogously, $\mathbf{B}_{k\leftarrow \ell}$ for $k \neq \ell$ relates to the transmission of information across groups. Hence we define the spillover from group ℓ to group k as $\mathcal{F}_{k\leftarrow \ell}^{(h)} = \mathbf{e}'_{m_k}\mathbf{B}_{k\leftarrow \ell}^{(h)}\mathbf{e}_{m_\ell}$ and the spillover to group k from group ℓ as $\mathcal{T}_{\ell\leftarrow k}^{(h)} = \mathbf{e}'_{m_\ell}\mathbf{B}_{\ell\leftarrow k}^{(h)}\mathbf{e}_{m_k}$. By collecting all these measures, we can define the h-step ahead block connectedness matrix of dimension $b \times b$ as

$$\underset{(b\times b)}{\mathbb{B}}^{(h)} = \begin{bmatrix} \mathcal{W}_{1\leftarrow -1}^{(h)} & \mathcal{F}_{1\leftarrow -2}^{(h)} & \cdots & \mathcal{F}_{1\leftarrow b}^{(h)} \\ \mathcal{F}_{2\leftarrow 1}^{(h)} & \mathcal{W}_{2\leftarrow 2}^{(h)} & \cdots & \mathcal{F}_{2\leftarrow b}^{(h)^{(h)}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{F}_{b\leftarrow 1}^{(h)} & \mathcal{F}_{b\leftarrow 2}^{(h)} & \cdots & \mathcal{W}_{b\leftarrow b}^{(h)} \end{bmatrix}. \tag{A.3}$$

Note that the dimension of this grouped matrix is $b^2 < K^2$, which implies a significant improvement on the FEVD interpretation in large models ease the processing constraints encountered in large models. It is now straightforward to develop aggregate connectedness measures at the group level.

Data description and preliminary analysis

Among the economic variables, we analyse the Industrial Production index, a business cycle indicator which measures monthly changes in the price-adjusted output of industry. The second variable is the index of the volume of Retail Trade, a business indicator which measures the monthly changes of the deflated turnover of Retail Trade both at the level of the European Union (EU) and of the Eurozone. Both the economic indices are provided in

their seasonally and calendar adjusted versions and normalized to 100 for the year 2015. As a measure of sentiment and expectations towards the future of economies, we select a third variable, i.e. the monthly Economic Sentiment index provided by the Directorate-General for Economic and Financial Affairs (DG ECFIN). The indicator is calculated on the basis of a selection of questions from six different sectors and different topics:

- Consumer: financial situation, general economic situation, price trends, unemployment;
- Industry: production, employment expectations, stocks of finished products and selling price;
- Services: business climate, evolution of demand, evolution of employment and selling prices;
- Financial services: business situation, evolution of demand and employment;
- Retail Trade: business situation, stocks of goods, orders placed with suppliers and firm's employment;
- Construction: trend of activity, order books, employment expectations, price expectations and factors limiting building activity.

Each indicator is the core of the surveys which leads to the construction of sectorial monthly confidence indices, aimed at reflecting the overall perceptions and expectations at the individual sector level in a one-dimensional index. Economic Sentiment is then formulated by aggregating information from the six confidence indices, obtaining a comprehensive measure which is able to track the overall economic activity. All of the economic variables are publicly available and provided by EUROSTAT[3].

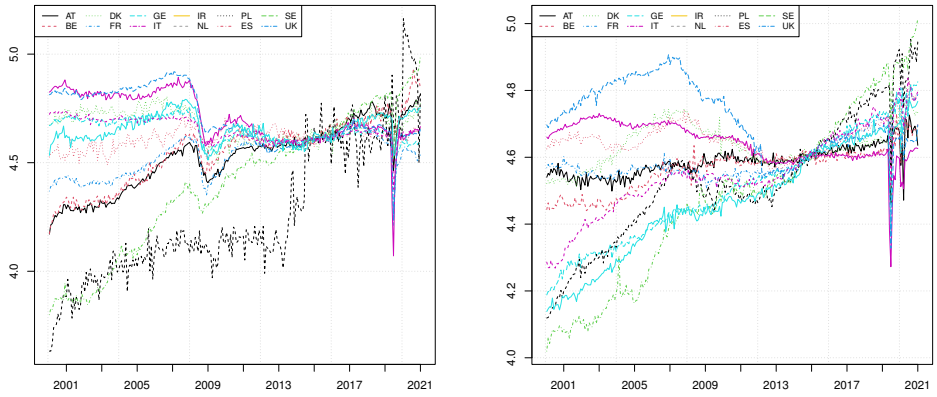
[3]Available at EUROSTAT: https://ec.europa.eu/eurostat/data/database

As far as the weights matrix used to build the foreign variables of the country models, we use the annual data on bilateral trade flows, which are made publicly available by the International Monetary Fund (IMF) [4]. We therefore define $\omega_{ij} = \omega_{ji}$ as the sum of the import and export of country i with country j.

Figure 26 shows the dynamics of the time series of Industrial Production, Retail Trade and Economic Sentiment of the set of 12 analyzed European countries over the whole sample period. It evidently appears how different are the time series in their temporal dynamics. As for the Industrial Production and Retail Trade indicators, countries have particularly different values in the first half of the time series, while starting from 2016 values tend to converge. Moreover, we can spot a clear drop along all the time series in 2020 (COVID-19 pandemic) and another significant one in 2009-2010, with major evidence on the sentiment index.

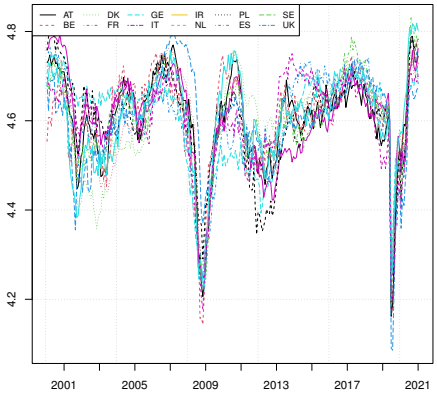
[4] Available at IMF: https://data.imf.org/?sk=9D6028D4-F14A-464C-A2F2-59B2CD424B85&sId=1409151240976

Figure 26: Time series of the indicators



(a) Industrial Production

(b) Retail Trade



(c) Economic sentiment

The figure shows the dynamics of the time series of Industrial Production, Retail Trade and Economic Sentiment of the set of 12 analyzed European countries expressed in logarithmic terms over the whole sample period.

Table 5: Weak exogeneity test

Country	p_i	q_i	C. R.	Crit. Val.	IND	RET	SENT
AT	1	1	2	F(2,250)	0,780	5,224*	3,855
BE	3	3	2	F(2,236)	0,478	0,265	0,577
DK	1	1	0	F(0,252)			
FR	3	2	2	F(2,239)	1,275	4,122	0,108
GE	2	2	0	F(0,245)			
IR	1	1	0	F(0,252)			
IT	3	3	1	F(1,237)	2,502	2,486	3,166
NL	1	1	0	F(0,252)			
PL	1	1	0	F(0,252)			
SP	1	2	1	F(1,247)	0,527	0,671	0,241
SE	1	1	1	F(1,251)	4,442	7,490*	0,996
UK	1	1	1	F(1,251)	6,681*	0,002	0,229

F-statistics for testing the weak exogeneity of the country-specific foreign variables. * denotes statistical significance at the 1% level. CR is the number of cointegrating relationship found in the single country model.

Weak exogeneity test

A first estimation attempt consists of using OLS to estimate the single country VARX models. However, it is well known that the main assumption underlying the estimation of the individual country VARX models is the weak exogeneity of $\mathbf{x}_{it}^{*}$ with respect to the long run parameters of the conditional model. This assumption might be in practice violated.

For this reason, we firstly estimate the single country VARX models by means of OLS, and then verify the validity of this assumption through the weak exogeneity test outlined in Harbo et al. (1998). The results are reported in Table A.1.

The weak exogeneity assumptions are rejected only in 3 cases out of 21. The first one pertains Industrial Production of UK, which is consistent with Dees et al. (2007), who found a rejection in the UK country model. The second and third ones are Retail Trade of Austria and Sweden. Given these premises and the suggestion of an anonymous Referee, we change our modelling strategy towards a more consistent estimation procedure outlined in the methodological section.

Global analysis

In order to measure the impacts of shocks on the real economy, the economic sentiment and their interrelationships across countries on a time-varying basis, we conduct a dynamic analysis. In other words, we set a rolling window w to estimate the Bayesian GVAR parameters and we obtain time varying estimates of the global forecast error variance decomposition connectedness measures. Our results are therefore obtained considering a Bayesian GVAR model based on twelve countries approximating VAR models estimated dynamically on a rolling window basis of six years and an H step ahead forecast horizon of 40 months.

As a preliminary analysis, we dynamically test for possible cointegrating relations among time series (in log-levels) through the Johansen Maximum Eigenvalue test. Figure 27 shows the proportion of significant cointegrating relationships with respect to the total number of possible cointegrating relationships present in the country models. Results clearly highlight that the number of cointegrating relationships tends to increase during crisis periods (see the Global Financial Crisis and the COVID-19 outbreak periods). This is equivalent to say that, in turbulent times, Industrial Production, Retail Trade and Economic Sentiment start to co-move more significantly. Interestingly, first evidence on the COVID-19 period supports the fact that the number of cointegrating relationships were immediately larger due the COVID-19 outbreak, if compared to those observed during the financial cri-

sis of 2008. In spite of this, the former had an instantaneous impact on the cointegrating relationships of the system variables, meaning that the shock seems to dissipate relatively quicker over time, while the latter had more persistent effects on time series co-movements.

Figure 29 shows the MST representation of the predictive spillover networks relative to the four periods of interest. It clearly appears a temporal dynamics in the backbone structure of the network. Evidence shows that the GFEVD network clusters are not constant and evolve dramatically over time - see Figure 30. We can see that the COVID-19 period shows the strongest links and the denser structure. Moreover, the two crisis periods are completely divergent: during the Global Financial Crisis the network is clustered and with few relevant links. During the COVID-19 period, instead, not only does the intensity of links increase, but also the number of clusters decrease, suggesting a higher degree of interconnectedness.

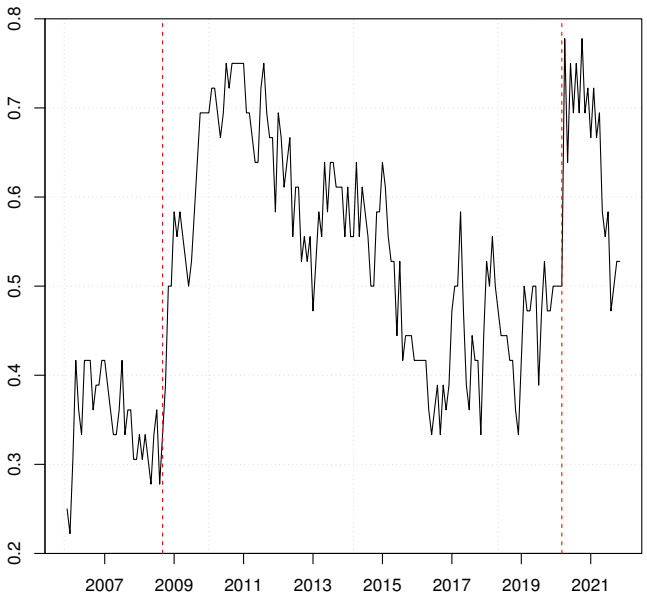
Country analysis

Figure 31 shows dependence-influence country relationships. The scatter plot is obtained by generating a Cartesian plane with the influence index on the x axis and the dependence index on the y axis. Both the indices are calculated from the country aggregation of the GFEVDs. Similarly to what observed in Figure 3, the first three periods are again comparable, despite some variability in the country specific positions in terms of dependence-influence. The real change occurs in phase 4, in which countries are basically equal in terms of dependence, allegedly due to the effect of the exogenous shock exerted by the pandemic outbreak.

Figure 28 shows the directed weighted GFEVD spillover network obtained from the country aggregation of the GFEVD during the four phases under analysis. In accordance with the results of the hub and authority algorithms, we find that Netherlands is a central node in the network in terms of shock

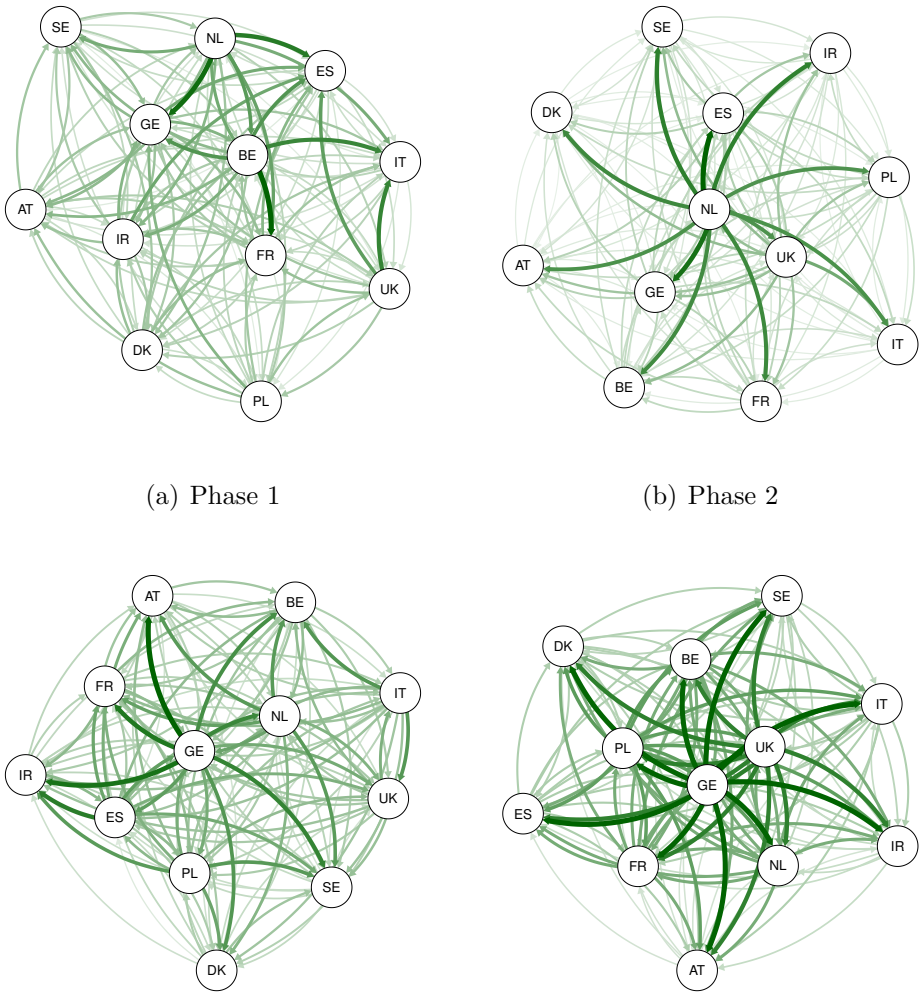
spreading power, especially during the Global Financial Crisis. Conversely, and still in line with the previous results, Germany appears to gain a pivotal role in the network in the aftermath of the Global Financial Crisis, which lasts even after the COVID-19 outbreak.

Figure 27: Fraction of significant cointegrating relationships



The figure shows the dynamics of total number of significant cointegrating relations in country models over the total number of possible cointegrating relationships (black line). The beginning of the Global Financial Crisis (September 2008) and of the COVID-19 outbreak (February 2020) are marked in red. Results refer to the Johansen Maximum Eigenvalue test for multiple cointegrating relationships considering a significance level of 10%.

Figure 28: Directed dynamic GFEVD country spillover network



(a) Phase 1

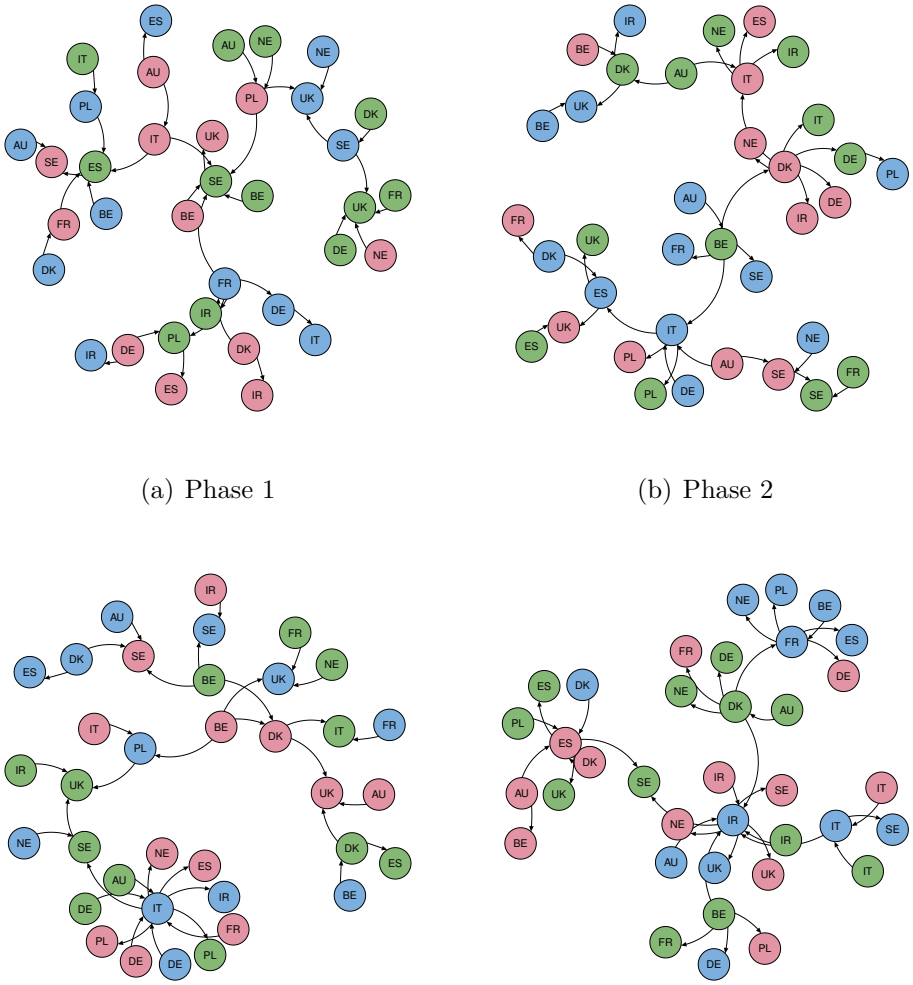
(b) Phase 2

(c) Phase 3

(d) Phase 4

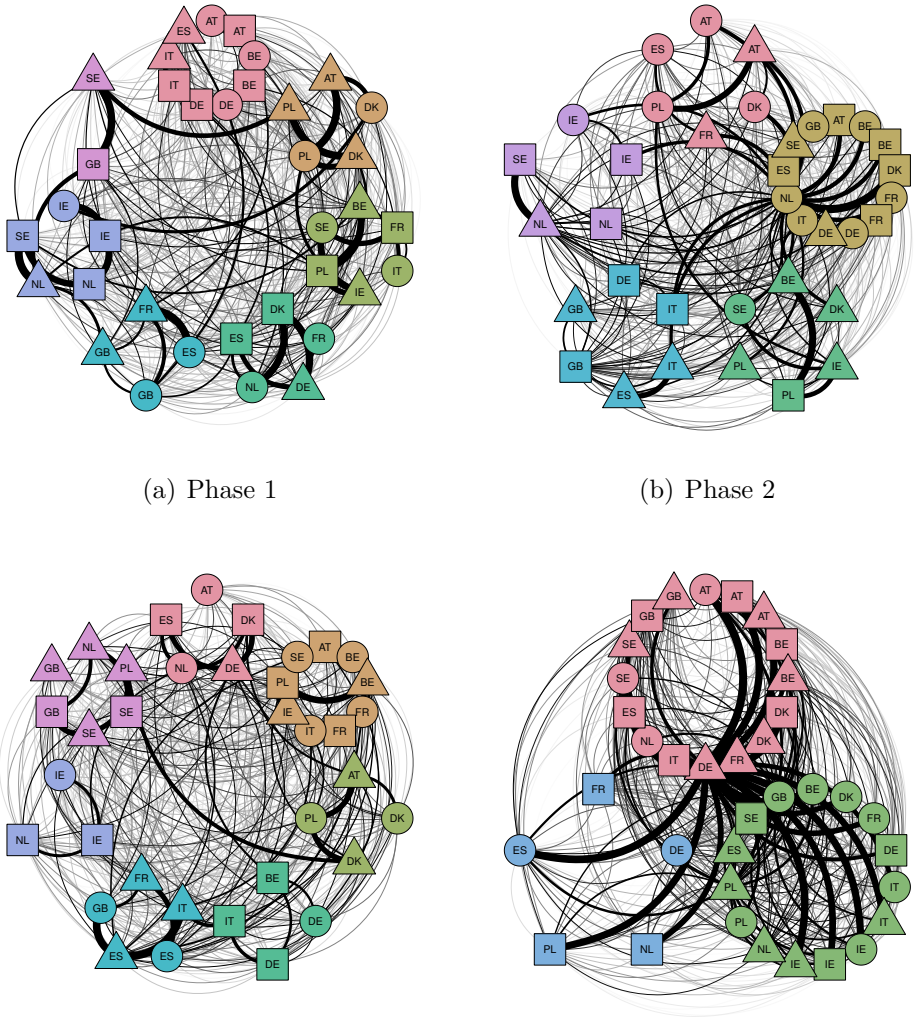
Directed GFEVD country spillover network. The figure shows the directed weighted GFEVD spillover network obtained from the country aggregation of the GFEVD phase by phase. Nodes represent countries, whereas links represent the magnitude of pairwise forecast error variance transmitted to others. Results are relative to the four sub-samples under analysis. Self-loops are omitted from the graphical representation.

Figure 29: Sub-sample global forecast error variance MSTs


ES
IT
AU
NE
NE
PL
PL
UK
AU
SE
ES
IT
UK
DK
SE
SE
BE
BE
UK
FR
FR
BE
DE
NE
DK
FR
DE
IR
PL
IR
DE
DK
IT
ES
IR
(a) Phase 1
IR
BE
NE
ES
DK
AU
IR
IT
UK
BE
IT
NE
DK
DE
PL
AU
DE
IR
FR
UK
DK
FR
BE
SE
ES
UK
IT
ES
NE
PL
AU
SE
FR
PL
DE
(b) Phase 2
IR
AU
SE
FR
ES
DK
SE
BE
UK
NE
IT
BE
DK
IT
FR
IR
PL
UK
UK
AU
NE
SE
DK
ES
AU
NE
ES
DE
IT
IR
PL
FR
DE
DE
PL
(c) Phase 3
NE
PL
BE
FR
ES
FR
DE
DE
ES
DK
NE
DK
AU
PL
ES
DK
SE
IR
SE
IT
AU
UK
NE
IR
IT
SE
BE
IR
AU
UK
UK
IT
BE
FR
DE
PL
(d) Phase 4

The figure shows the MST representation derived from the generalized forecast error variance decomposition (GFEVD) relative to the four periods of interest. Network nodes represent system variables, whereas edges represent the directional contribution of each variable to the forecast error variance decomposition of the others. Colours stand for Industrial Production (red), Retail Trade (green) and Economic Sentiment (blue). Self-loops are omitted.

Figure 30: Sub-sample global forecast error variance clustering



(a) Phase 1

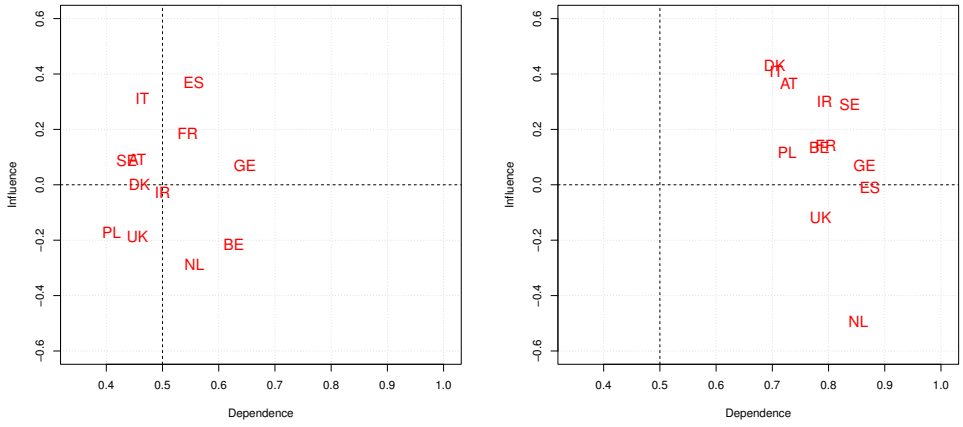
(b) Phase 2

(c) Phase 3

(d) Phase 4

The figure shows the clustered generalised forecast error variance decomposition (GFEVD) network structure of real economy and economic sentiment variables, where groups are classified according to the Louvain clustering algorithm. Squares: Industrial Production, triangles: Retail Trade, circles: Economic Sentiment.

Figure 31: Dependence-influence country relationships



(a) Phase 1

(b) Phase 2

(c) Phase 3

(d) Phase 4

The figure shows a scatter plot obtained by generating a Cartesian plane with the influence index on the x axis and the dependence index on the y axis. Both the indices are calculated from the country aggregation of the rolling GFEVDs.

A.2 Technical details of Chapter 3

Tensor operations and the Tucker product

A tensor is a multidimensional array, whose order expresses the number of dimensions, also known as ways or modes To avoid confusion, we use the term way to express the dimension of a tensor, given that mode is already used to express the maximum value of a given distribution. More formally, an N-th way tensor is an N dimensional array $\mathcal{X} \in \mathbb{R}^{I_1 \times ... \times I_N}$ with entries $\mathcal{X}_{i_1 ... i_N}$ with $i_n = 1, ..., I_n$ and $n = 1, ..., N$.

Vectors are tensors of order one (denoted by boldface lowercase letters, e.g., $\mathbf{x}$) whereas matrices are tensors of order two (denoted by boldface capital letters, e.g., $\mathbf{X}$).

Tensor norm and inner product

The Frobenius norm of a tensor $\mathcal{X}$ is the square root of the squared sum of all its elements, ie.

$$||\mathcal{X}||_F = \sqrt{\sum_{j_1=1}^{J_1} ... \sum_{j_N=1}^{J_N} x_{j_1 ... j_N}^2} = \sqrt{\text{vec}(\mathcal{X})' \text{vec}(\mathcal{X})} \tag{A.4}$$

which is analogous to the Frobenious norm of a matrix.

The inner-product of two tensors of the same dimension $\mathcal{X}, \mathcal{Y} \in \mathbb{R}^{J_1, ..., J_N}$ is the sum of the product of their entries:

$$\langle \mathcal{X}, \mathcal{Y} \rangle = \sum_{j_1=1}^{J_N} .... \sum_{j_N=1}^{J_N} x_{j_1 ... j_N} y_{j_1 ... j_N} = \text{vec}(\mathcal{X})' \text{vec}(\mathcal{Y}) \tag{A.5}$$

Matricization and Tucker product

The process of reordering the elements of an N-way array into a matrix is called matricization. The n-th way matricization of $\mathcal{X}$ is denoted by $\mathbf{X}^n =$

$\text{mat}_k(\mathcal{X})$, and is obtained by reshaping the elements of the original tensor so that the resulting matrix is of dimension $[I_n \times \prod_{j \neq n} I_j]$. The special case of contemporaneous matricization along all the ways of a tensor is called vectorization, which is analogous to the vectorization of a matrix:

$$\mathbf{x} = \text{vec}(\mathcal{X}) = \text{mat}_{1,...,N}(\mathcal{X}) \tag{A.6}$$

Given the matrices $\mathbf{B}_1, ..., \mathbf{B}_N$ with $\mathbf{B}_n \in \mathbb{R}^{i_n \times j_n}$, a map from the space of $\mathcal{X}$ to the space generated by the rows of $\mathbf{B}_n$ ($\mathbb{R}^{i_1 \times ... \times i_N}$) is made by first obtaining $\mathbf{x}$, then computing:

$$\mathbf{m} = (\mathbf{B}_N \otimes .... \otimes \mathbf{B}_1)\mathbf{x} \tag{A.7}$$

and eventually forming an $[i_1 \times ... \times i_N]$ dimensional array $\mathcal{M}$ from $\mathbf{m}$. This transformation between the tensor $\mathcal{X}$ and the list $\mathbf{B} = \{\mathbf{B}_1, ..., \mathbf{B}_N\}$ is known as the Tucker product (Tucker, 1966), and can be written as:

$$\mathcal{M} = \mathcal{X} \bar{\times} \{\mathbf{B}_1, ...\mathbf{B}_N\} \tag{A.8}$$

It is worth noting that the matricization operator connects the multidimensional Tucker product to the well known matrix multiplication, facilitating both understanding and computation of the former. In fact, by applying the n-th way matricization to both sides of equation (A.8) we obtain the equivalent formulation:

$$\mathbf{M}^n = \mathbf{B}_n \mathbf{X}^n \mathbf{B}^{'}_{-n} \tag{A.9}$$

where $\mathbf{B}_{-n} = (\mathbf{B}_N \otimes ... \otimes \mathbf{B}_{n+1} \otimes \mathbf{B}_{n-1} \otimes ... \otimes \mathbf{B}_1)$. By repeating the operation for $n = 1, ..., N$, it emerges that the Tucker product can be expressed as a series of N matrix reshaping and multiplications. Matricization and vectorization applied to the Tucker product give raise to the following set of equivalences:

$$\begin{aligned} \mathcal{M} &= \mathcal{X} \bar{\times} \{\mathbf{B}_1, ..., \mathbf{B}_N\} \\ \mathbf{M}^n &= \mathbf{B}_n \mathbf{X}^n \mathbf{B}^{'}_{-n} \\ \mathbf{m} &= (\mathbf{B}_N \otimes .... \otimes \mathbf{B}_1)\mathbf{x} \end{aligned} \tag{A.10}$$

The VAR as well as the MAR equivalent form of a TAR can be easily derived with the abovementioned tools.

Conditional Posterior Distribution

The conditional posterior of each $\gamma_{i,k}$ is:

$$\pi(\gamma_{i,k}|-) \sim \mathcal{B}\text{er}(\bar{\theta}_{i,k}) \tag{A.11}$$

where

$$\bar{\theta}_{i,k} = \frac{\theta_k \mathcal{N}(\phi_{i,k}, \tau_1)}{\theta_k \mathcal{N}(\phi_{i,k}, \tau_1) + (1-\theta_k)\mathcal{N}(\phi_{i,k}, \tau_0)} \tag{A.12}$$

The conditional posterior of each $\boldsymbol{\phi}_k$ is:

$$\pi(\boldsymbol{\phi}_k|-) \sim \mathcal{N}(\bar{\mu}_k, \bar{\Omega}_k) \tag{A.13}$$

where

$$\begin{aligned} \bar{\Omega}_k &= (\tilde{\mathbf{X}}^k \tilde{\mathbf{X}}^{k'} \otimes \boldsymbol{\Sigma}_k^{-1} + \mathcal{V}_k^{-1})^{-1} \\ \bar{\mu}_k &= \bar{\Omega}_k (\tilde{\mathbf{X}}^k \tilde{\mathbf{X}}^{k'} \otimes \boldsymbol{\Sigma}_k^{-1} \hat{\boldsymbol{\phi}}_k) \end{aligned} \tag{A.14}$$

where $\hat{\boldsymbol{\Phi}}_k = (\tilde{\mathbf{Y}}^k \tilde{\mathbf{X}}^{k'})(\tilde{\mathbf{X}}^k \tilde{\mathbf{X}}^{k'})^{-1}$ and $\hat{\boldsymbol{\phi}}_k = \text{vec}(\hat{\boldsymbol{\Phi}}_k)$.

The conditional posterior of each $\boldsymbol{\Sigma}_k$ is:

$$\pi(\boldsymbol{\Sigma}_k|-) \sim \mathcal{W}^{-1}(\bar{\Omega}_k, n_{2,k} + \nu_k) \tag{A.15}$$

where $\bar{\Omega}_k = (\tilde{\mathbf{Y}}^k - \hat{\boldsymbol{\Phi}}_k \tilde{\mathbf{X}}^k)(\tilde{\mathbf{Y}}^k - \hat{\boldsymbol{\Phi}}_k \tilde{\mathbf{X}}^k)' + \xi\Omega_k$.

The conditional posterior of each θ_k is:

$$\pi(\theta_k|-) \sim \mathcal{B}\text{eta}(\bar{\alpha}_k, \bar{\beta}_k) \tag{A.16}$$

where $\bar{\alpha}_k = |\boldsymbol{\gamma}_k| + \alpha_k$ and $\bar{\beta}_k = N_k + \beta_k + \alpha_k$.

The conditional posterior for ξ is:

$$\pi(\xi|-) \sim \mathcal{G}a(\bar{\nu}_1, \bar{\nu}_2) \tag{A.17}$$

where $\bar{\nu}_1 = \frac{1}{2}\sum_k d_k \nu_k + \nu_1$ and $\bar{\nu}_2 = \frac{1}{2}\sum_k \Omega_k \boldsymbol{\Sigma}_k^{-1} + \nu_2$.

E-M steps

For each k, the E-steps proceeds by computing the conditional expectations of $v_{i,k}^{-1} \in \mathcal{V}_k^{-1}$ in $\mathcal{Q}_{1,k}(\cdot)$ and of $\gamma_{i,k}$ for $|\boldsymbol{\gamma}_k|$ in $\mathcal{Q}_{3,k}(\cdot)$. Consider the latter first. At the j-th step we have:

$$\begin{aligned} \mathrm{E}_{\boldsymbol{\gamma}_k|\cdot}(\gamma_{i,k}) &= \mathrm{P}(\gamma_{i,k} = 1|\boldsymbol{\phi}_k^{[j-1]}, \theta_k^{[j-1]}) \\ &= \frac{\theta_k^{[j-1]}\mathcal{N}(\phi_{i,k}|0,\tau_1)}{\theta_k^{[j-1]]}\mathcal{N}(\phi_{i,k}|0,\tau_1) + (1-\theta_k^{[j-1]})\mathcal{N}(\phi_{i,k}|0,\tau_0)} \end{aligned} \tag{A.18}$$

The E-step for the former is:

$$\begin{aligned} \mathrm{E}_{\boldsymbol{\gamma}_k|\cdot}(v_{i,k}^{-1}) = \mathrm{E}_{\boldsymbol{\gamma}_k|\cdot}[\tau_0(1-\gamma_{i,k}) + \tau_1\gamma_{i,k}]^{-1} = \frac{1 - \mathrm{P}(\gamma_{i,k} = 1|\boldsymbol{\phi}_k^{[j-1]}, \theta_k^{[j-1]})}{\tau_0} \\ + \frac{\mathrm{P}(\gamma_{i,k} = 1|\boldsymbol{\phi}_k^{[j-1]}, \theta_k^{[j-1]})}{\tau_1} \end{aligned} \tag{A.19}$$

The maximization steps are:

$$\boldsymbol{\phi}_k = [\mathcal{V}^{-1} + (\tilde{\mathbf{X}}^{k'}\tilde{\mathbf{X}}^k \otimes \boldsymbol{\Sigma}_k^{-1})]^{-1}[(\tilde{\mathbf{X}}^{k'}\tilde{\mathbf{X}}^k \otimes \boldsymbol{\Sigma}^{-1})\hat{\boldsymbol{\phi}}_k] \tag{A.20}$$

$$\boldsymbol{\Sigma}_k = \frac{(\tilde{\mathbf{Y}}^k - \boldsymbol{\Phi}_k\tilde{\mathbf{X}}^k)(\tilde{\mathbf{Y}}^k - \boldsymbol{\Phi}_k\tilde{\mathbf{X}}^k)') + \xi\Omega_k}{J_k + J_{-k} + \nu_k + 1} \tag{A.21}$$

$$\theta_k = \frac{|\boldsymbol{\gamma}_k| + \alpha_k - 1}{n_k + \alpha_k + \beta_k + -2} \tag{A.22}$$

$$\xi = \frac{\frac{1}{2}\sum_k(J_k\nu_k) + \eta_1 - 1}{\frac{1}{2}\sum_k \mathrm{tr}(\Omega_k\boldsymbol{\Sigma}_k^{-1}) + \eta_2} \tag{A.23}$$

Additional Simulation Results

Competing models

In compact form, a PVAR mode can be rewritten as:

$$\mathbf{Y} = \mathbf{B}\mathbf{X} + \mathbf{U}, \tag{A.24}$$

or

$$\mathbf{y} = (\mathbf{X}^{'} \otimes \mathbf{I}_{GN})\boldsymbol{\beta} + \boldsymbol{u}, \tag{A.25}$$

where $\mathbf{Y} = [\mathbf{y}_{P+1}, ..., \mathbf{y}_T]$ and the coefficient matrix $\mathbf{B} = [\boldsymbol{\Phi}_1, ..., \boldsymbol{\Phi}_P]$ is of dimension $GN \times GNP$. $\mathbf{X} = [\mathbf{X}_P, ..., \mathbf{X}_{T-1}]$, with $\mathbf{X}_t = [\mathbf{y}_{t-1}, ..., \mathbf{y}_{t-P}]$. Equation (A.25) is the vectorized form, where $\mathbf{y} = \text{vec}(\mathbf{Y})$, $\mathbf{u} = \text{vec}(\mathbf{U})$ and $\boldsymbol{\beta} = \text{vec}(\mathbf{B})$.

We consider the following competing models:

1. **CC**: Canova and Ciccarelli (2009, 2013) use a factorization approach of the parameters such that they can be divided into common, country-specific, and variable-specific factors. They specify the model in a hierarchical structure:

$$\begin{aligned} \boldsymbol{\beta}|\mathbf{F} &\sim \mathcal{N}(\Lambda\mathbf{F}, \boldsymbol{\Sigma} \otimes \mathbf{I}_{GN}), \\ \mathbf{F} &\sim \mathcal{N}(0, \underline{c}\mathbf{F}), \end{aligned} \tag{A.26}$$

   where Λ is an $GN \times f$ matrix of loadings and $\mathbf{F}$ is an f-dimensional vector of factors where $f < GN$. The number of factors are, respectively, N for coefficients of each country and G for coefficients of each variable, and one common factor for all coefficients. There is only one hyperparameter to set is $\underline{c}$, related to the prior variance of the factors.

2. **SSVS**: George et al. (2008) specify a prior whereby each coefficient of $\mathbf{B}$ is drawn from a mixture of two normal distributions: the former with a small variance aiming at shrinking the coefficient towards 0 and the latter with a relatively large one. The higher the magnitude of $\mathbf{B}_{ij}$, the higher is the probability that it will be drawn from the second distribution, and viceversa.

$$\begin{aligned} \boldsymbol{\beta}_k|\gamma_k &\sim (1-\gamma_k)\mathcal{N}(0, \underline{\tau}_1^2) + \gamma_k\mathcal{N}(0, \underline{\tau}_2^2), \\ \gamma_k &\sim \mathcal{B}\mathbf{er}(\underline{\pi}_k), \end{aligned} \tag{A.27}$$

   with $k = 1, ..., G^2N^2P$.

3. **SSSS**: This algorithm designed by Koop and Korobilis (2016), who build on George et al. (2008) but taking in into account for panel restrictions. They specify two types of restrictions: dynamic interdependencies (DI) and cross-sectional homogeneity (CSH).

   The DI works on off-diagonal blocks. Let $\mathbf{B}_{ij} \in \mathbf{B}$ be the $G \times G$ block embodying the parameters of country j-th on country i-th equations. The prior has the following form:

$$\begin{aligned} \text{vec}(\mathbf{B}_{ij})|\gamma_{ij}^{DI} &\sim (1-\gamma_{ij}^{DI})\mathcal{N}(0, \underline{\tau}_1^2 \mathbf{I}_G{}^2) + \gamma_{ij}^{DI}\mathcal{N}(0, \underline{\tau}_2^2 \mathbf{I}_G{}^2), \\ \gamma_{ij}^{DI}|\pi_{ij}^{DI} &\sim \mathcal{B}\mathbf{er}(\pi_{ij}^{DI}), \forall j \neq i, \\ \pi_{ij}^{DI} &\sim \mathcal{B}\mathbf{eta}(1, \underline{\phi}), \end{aligned} \tag{A.28}$$

   whereas the CSH prior works on the main block diagonal of $\mathbf{B}$. The prior reads as:

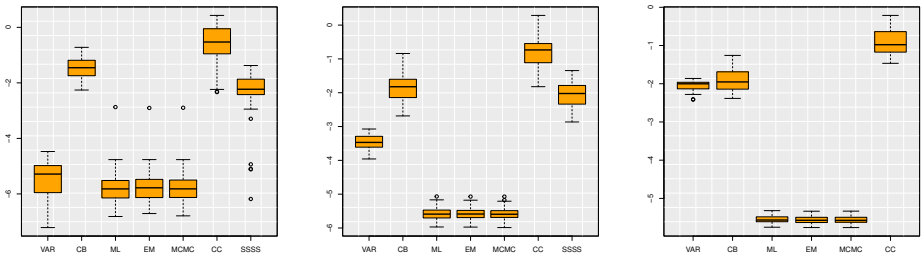
$$\begin{aligned} \text{vec}(\mathbf{B}_{ii})|\gamma_{ij}^{CSH} &\sim (1-\gamma_{ij}^{CSH})\mathcal{N}(\mathbf{B}_{jj}, \underline{\xi}_1^2 \mathbf{I}_G^2) + \gamma_{ij}^{CSH}\mathcal{N}(\mathbf{B}_{jj}, \underline{\xi}_2^2 \mathbf{I}_G^2), \\ \gamma_{ij}^{CSH}|\pi_{ij}^{CSH} &\sim \mathcal{B}\mathbf{er}(\pi_{ij}^{CSH}), \quad \forall j \neq i, \\ \pi_{ij}^{CSH} &\sim \mathcal{B}\mathbf{eta}(1, \underline{\phi}). \end{aligned} \tag{A.29}$$

Hyper/regularization parameter tuning

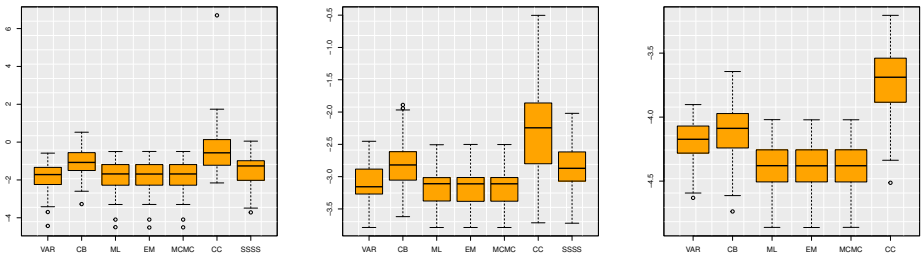
- **CC**: We set $\underline{c} = 4$.
- **SSVS**: We set $\underline{\tau}_1^2 = 0.01$, $\underline{\tau}_2^2 = 4$ and $\underline{\pi}_k = 0.5$.
- **SSSS** We set $\underline{\tau}_1^2, \underline{\xi}_1^2 = 0.01$, $\underline{\tau}_2^2, \underline{\xi}_2^2 = 4$, $\pi_{ij}^{DI}, \pi_{ij}^{CSH} = 0.5$ and $\underline{\pi}_k = 1$.

Additional results

Figure 32: Average MSE and 1 step ahead forecasting performance



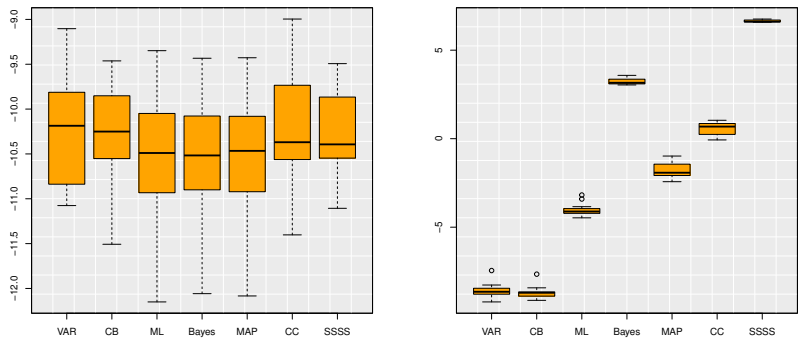
(a) G=2, N=3 (b) G=4, N=6 (c) G=8, N=10



(d) G=2, N=3 (e) G=4, N=6 (f) G=8, N=10

Average estimation error (MSE) (a,b,c) and 1 step ahead forecasting performance (MSFE) (d,e,f) over 50 repetitions of each model for $(G, N) = (2, 3), (4, 6), (8, 10)$ with $T = 500$.

Figure 33: One step ahead MSFE and computational time



(a) G=2, N=3 (b) G=4, N=6

One step ahead MSFE (a) and computational time (b) of each model over different rolling windows spanning from 1995Q1 (T = 60) to 2019Q2 (T = 120).

A.3 Technical details of Chapter 4

Static connectedness

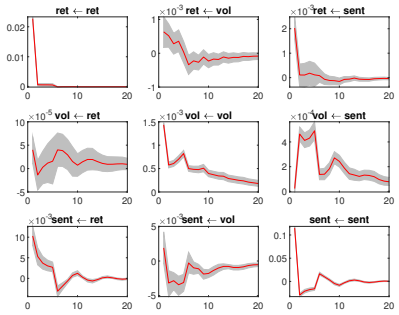
Given the MA(∞) representation of the MAR in vectorized form - see equation (4.5), we can perform dynamic analysis via impulse response functions. In particular, in Figure 34 we illustrate the generalized impulse response functions, for $H = 1, ..., 20$ relative to what we define a "global shock". A global shock is defined as such as it does not originate from a particular asset, but is common to the set of securities as a whole. In other words, we study shocks arising from endogenous variables of the system, defined as a weighted average of variable specific shocks across all the indicators of one asset (or across one indicator of all assets) in the autoregressive model.

The Generalized Impulse Response (GIRF) function (Koop et al., 1996) in the case of a one standard error global shock equally distributed across the assets at time t, considering a forecast horizon H, can be written as:

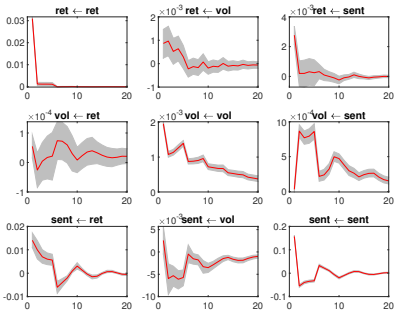
$$\mathcal{GIRF}(H) = \frac{\mathbf{\Gamma}_H(\mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1)\mathfrak{e}}{\sqrt{\mathfrak{e}'(\mathbf{\Sigma}_2 \otimes \mathbf{\Sigma}_1)\mathfrak{e}}}, \tag{A.30}$$

where Γ_H is the H step matrix of dynamic multipliers and $\mathfrak{e}$ is a $GN \times 1$ vector of shocks such that $\sum_i \mathfrak{e}_i = 1$. We illustrate the resulting impulse responses for each cryptocurrency associated to a positive 1 standard deviation shock in Figure 34.

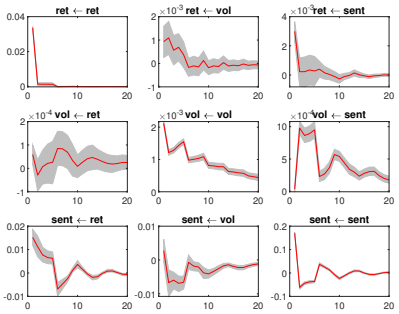
Forecasting experiment



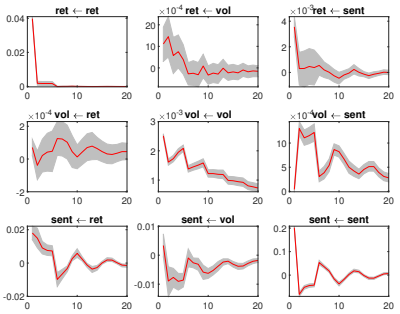
(c) Bitcoin



(d) Ethereum



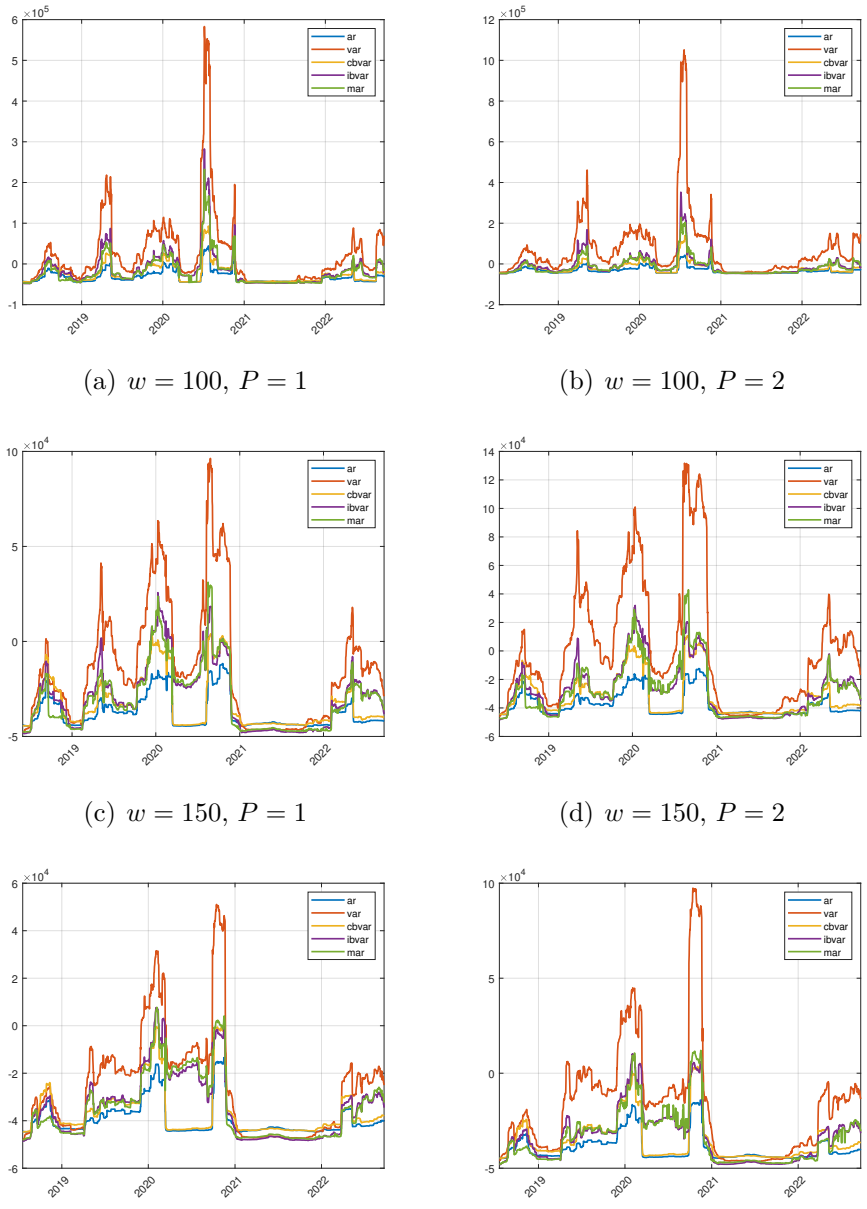
(e) Litecoin



(f) Ripple

Figure 34: Generalized impulse response functions for the four indicators.

Figure 35: Rolling BIC



(a) $w = 100$, $P = 1$

(b) $w = 100$, $P = 2$

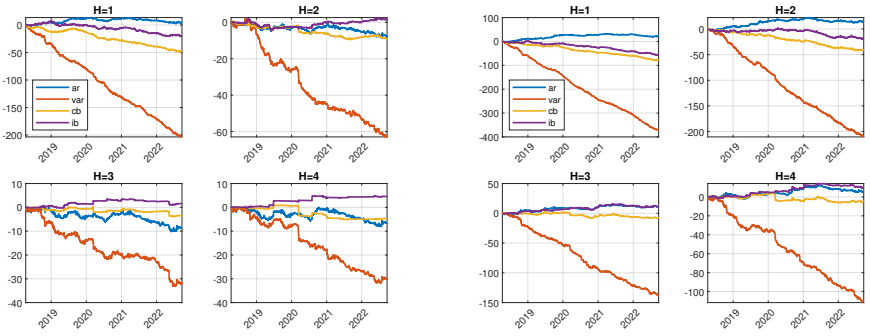
(c) $w = 150$, $P = 1$

(d) $w = 150$, $P = 2$

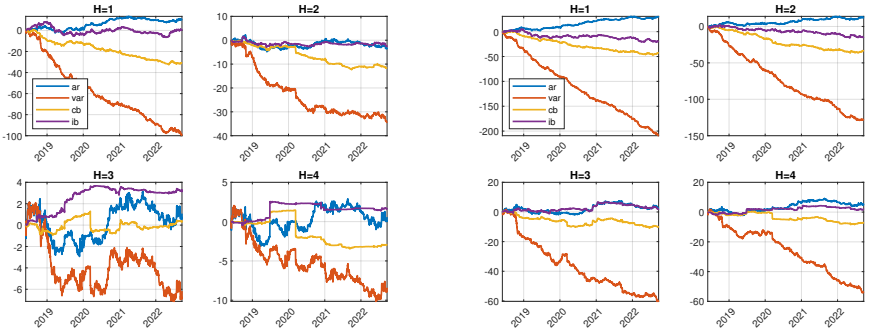
(e) $w = 200$, $P = 1$

(f) $w = 200$, $P = 2$

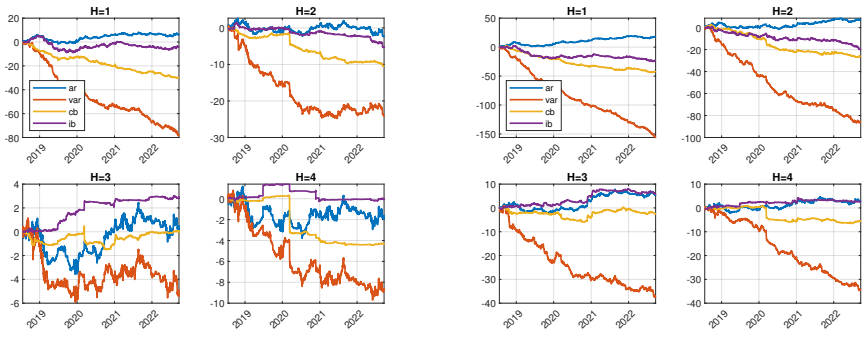
Rolling Bayes-Schwarz Information Criterion (BIC) for the five models considered.



(g) $w = 100$, $P = 1$ (h) $w = 100$, $P = 2$

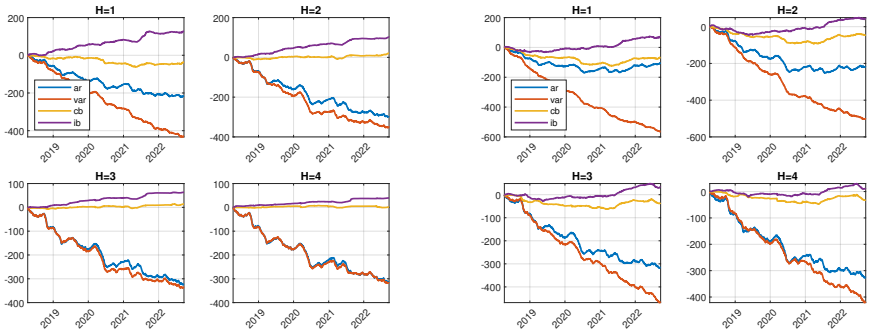


(i) $w = 150$, $P = 1$ (j) $w = 150$, $P = 2$



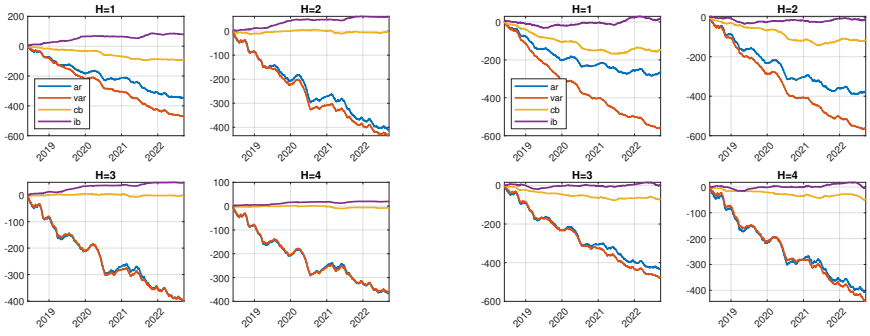
(k) $w = 200$, $P = 1$ (l) $w = 200$, $P = 2$

Figure 36: Rolling cumulative logarithm of the RMSFE for the return indicator and the five different models.



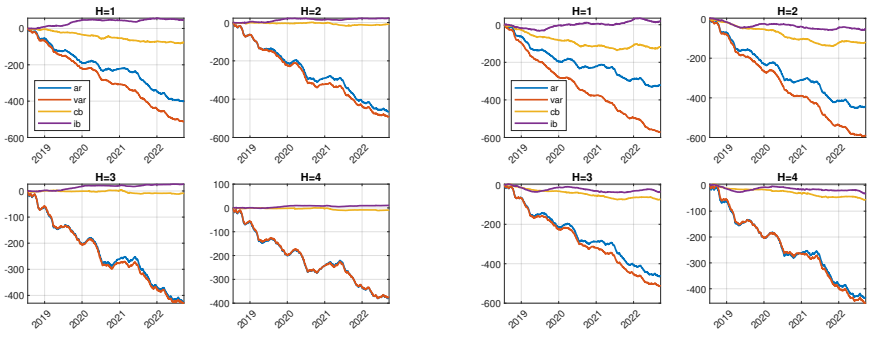
(a) $w = 100$, $P = 1$

(b) $w = 100$, $P = 2$



(c) $w = 150$, $P = 1$

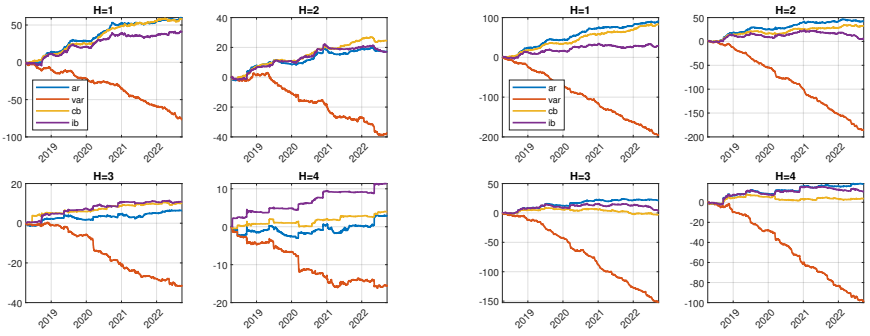
(d) $w = 150$, $P = 2$



(e) $w = 200$, $P = 1$

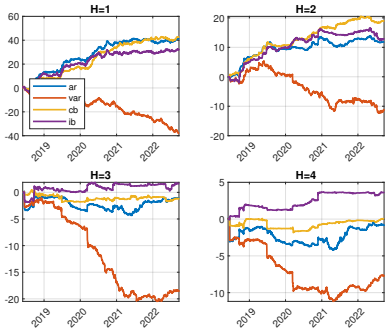
(f) $w = 200$, $P = 2$

Figure 37: Rolling cumulative logarithm of the RMSFE for the volatility indicator and the five different models.

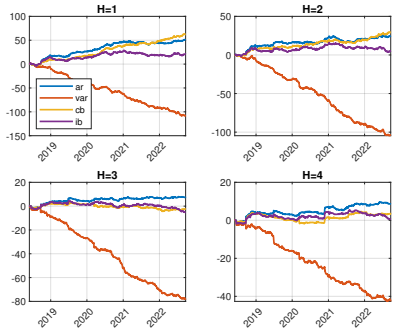


(a) $w = 100$, $P = 1$

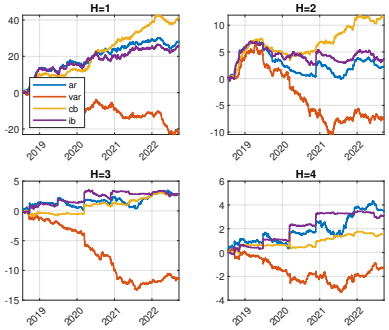
(b) $w = 100$, $P = 2$



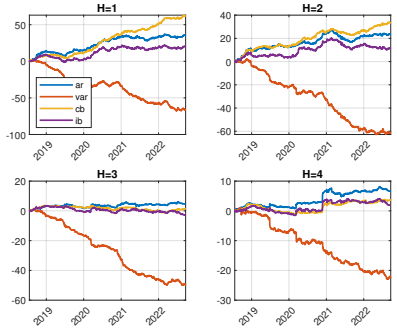
(c) $w = 150$, $P = 1$



(d) $w = 150$, $P = 2$



(e) $w = 200$, $P = 1$



(f) $w = 200$, $P = 2$

Figure 38: Rolling cumulative logarithm of the RMSFE for the sentiment indicator and the five different models.

Table 6: Diebold Mariano test P-value (MAR-AR)

	btc	eth	ltc	xrp
ret	0.133	0.059	0.055	0.010
vol	< 0.01	0.014	< 0.01	0.489
sent	0.285	0.370	< 0.01	< 0.01

FH 1

	btc	eth	ltc	xrp
ret	0.066	0.036	0.053	0.124
vol	0.241	0.047	< 0.01	0.721
sent	0.917	0.402	0.090	< 0.01

FH 2

	btc	eth	ltc	xrp
ret	0.813	0.070	0.373	0.136
vol	0.303	0.013	< 0.01	0.178
sent	0.499	0.254	0.131	< 0.01

FH 3

	btc	eth	ltc	xrp
ret	0.276	0.175	0.479	0.296
vol	0.053	< 0.01	< 0.01	0.283
sent	0.645	0.189	0.374	< 0.01

FH 4

P-value of the Diebold-Mariano test for each variable. The null hypothesis is that MAR and AR have the same forecast accuracy. Setting: $w = 150$, $P = 2$.

Table 7: Diebold Mariano test P-value (MAR-VAR)

	btc	eth	ltc	xrp
ret	< 0.01	< 0.01	< 0.01	< 0.01
vol	< 0.01	< 0.01	< 0.01	< 0.01
sent	< 0.01	< 0.01	< 0.01	< 0.01

FH 1

	btc	eth	ltc	xrp
ret	< 0.01	< 0.01	< 0.01	< 0.01
vol	< 0.01	< 0.01	0.017	< 0.01
sent	< 0.01	< 0.01	< 0.01	< 0.01

FH 2

	btc	eth	ltc	xrp
ret	< 0.01	< 0.01	0.016	< 0.01
vol	0.103	0.336	0.687	< 0.01
sent	< 0.01	< 0.01	< 0.01	< 0.01

FH 3

	btc	eth	ltc	xrp
ret	< 0.01	< 0.01	0.025	< 0.01
vol	0.972	0.667	0.600	< 0.01
sent	< 0.01	< 0.01	0.036	< 0.01

FH 4

P-value of the Diebold-Mariano test for each variable. The null hypothesis is that MAR and VAR have the same forecast accuracy. Setting $w = 150$, $P = 2$.

MCMC convergence

We have carried out a convergence diagnostics based on the LeSage' Econometrics toolbox [5] to evaluate the convergence of the MCMC algorithm in the full sample exercise. In particular, we rely on the diagnostic of Raftery and Lewis (1970).

Raftery and Lewis provide an answer regarding how long to monitor the chain in order to reach a predermined posterior summary accuracy. Specifically, given a posterior quantile q of interest, an acceptable tolerance r of the probability p of being within the interval $(q - r, q + r)$, it provides an estimates of the number of iterations (MC), the number of burn-in (BI) and the thinning period (TH) necessary to satisfy the specified conditions regarding quantile q. Also, the number of draws that would be needed if the draws represented an i.i.d. chain are given (MCmin). Finally, also the I-statistic (I-stat), which is given by $MC/MCmin$, is reported. It provides evidence of convergence problems if its values exceeds 5.

We compute these measures for some randomly chosen entries of the estimated coefficient matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$, related to the full sample experiment, respectively. The results are reported in Table 8.

As expected, results are coherent between the two set of coefficients. The thinning estimate provided by the second and seventh columns is 1, which is consistent with the lack of autocorrelation in the sequence of draws. Regarding the number of burn in, it is suggested that only 2 draws for $\mathbf{A}$ and 3 for $\mathbf{B}$ are required, which are pretty small. We remark that our starting values are represented by the mode of the posterior distribution, estimated via the MAP, which are supposed to be nearby the true values. The fourth and eight columns, report the number of draws needed to achieve the desired level of accuracy. In both cases the value is far below 5000, which corresponds to the number of draws we employed, confirming the overall efficieny

[5] Available at: https://www.spatial-econometrics.com/

Table 8: Raftery and Lewis diagnostics

	BI	TH	MC	MCmin	I-stat		BI	TH	MC	MCmin	I-stat
$\mathbf{A}_{3,2}$	1	3	4129	3746	1.102	$\mathbf{B}_{3,4}$	1	2	3866	3746	1.032
$\mathbf{A}_{2,4}$	1	3	4129	3746	1.102	$\mathbf{B}_{2,5}$	1	2	3866	3746	1.032
$\mathbf{A}_{3,4}$	1	3	4129	3746	1.102	$\mathbf{B}_{2,7}$	1	2	3866	3746	1.032
$\mathbf{A}_{2,7}$	1	3	4129	3746	1.102	$\mathbf{B}_{4,9}$	1	2	3866	3746	1.032
$\mathbf{A}_{3,7}$	1	3	4129	3746	1.102	$\mathbf{B}_{2,13}$	1	2	3866	3746	1.032
$\mathbf{A}_{1,10}$	1	3	4129	3746	1.102	$\mathbf{B}_{1,14}$	1	2	3866	3746	1.032
$\mathbf{A}_{1,11}$	1	3	4129	3746	1.102	$\mathbf{B}_{2,14}$	1	2	3866	3746	1.032
$\mathbf{A}_{2,12}$	1	3	4129	3746	1.102	$\mathbf{B}_{4,14}$	1	2	3866	3746	1.032
$\mathbf{A}_{1,12}$	1	3	4129	3746	1.102	$\mathbf{B}_{1,16}$	1	2	3866	3746	1.032
$\mathbf{A}_{3,12}$	1	3	4129	3746	1.102	$\mathbf{B}_{3,16}$	1	2	3866	3746	1.032

Raftery and Lewis diagnostics for ten randomly selected entries of $\hat{\mathbf{A}}$ (columns 1:6) and $\hat{\mathbf{B}}$ (columns 7:12). The column stand for the number of burn-in (BI) and the thinning period (TH), the number of Monte Carlo draws (MC), the minimum number of draws (MCmin), I-statistic (I-stat)

of the estimation process.