Ca' Foscari University of Venice

Department of Economics

Working Paper

Monica Billio
Massimiliano Caporin
Lorenzo Frattarolo
Loriana Pelizzon

Networks in risk spillovers: A multivariate GARCH perspective

ISSN: 1827-3580
No. 16/WP/2020

Working Papers
Department of Economics
Ca' Foscari University of Venice
No. 16/WP/2020
ISSN 1827-3580

Networks in risk spillovers: A multivariate GARCH perspective

Monica Billio
Ca' Foscari University of Venice

Massimiliano Caporin
Department of Statistical Sciences, University of Padua

Lorenzo Frattarolo
Ca' Foscari University of Venice

Loriana Pelizzon
SAFE-Goethe University Frankfurt (Germany); Ca' Foscari University of Venice

Abstract
We propose a spatiotemporal approach for modeling risk spillovers using time-varying proximity matrices based on observable financial networks and introduce a new bilateral Multivariate GARCH speci_cation. We study covariance stationarity and identification of the model, develop the quasi-maximum-likelihood estimator and analyze its consistency and asymptotic normality. We show how to isolate risk channels and we discuss how to compute target exposure able to reduce system variance. An empirical analysis on Euroarea bond data shows that Italy and Ireland are key players in spreading risk, France and Portugal are major risk receivers, and we uncover Spain's non-trivial role as risk middleman.

Keywords
Spatial GARCH, network, risk spillover, financial spillover

JEL Codes
C58, G10

Address for correspondence:
Massimiliano Caporin
Department of Statistical Sciences
University of Padua
Via C. Battisti 241
35121, Padova - Italy
E-mail: massimiliano.caporin@unipd.it

The Working Paper Series
is available only on line
(http://www.unive.it/pag/16882/)
For editorial correspondence, please contact:
wp.dse@unive.it

Department of Economics
Ca' Foscari University of Venice
Cannaregio 873, Fondamenta San Giobbe
30121 Venice Italy
Fax: ++39 041 2349210

Networks in risk spillovers: A multivariate GARCH perspective*

Monica Billio[†] **Massimiliano Caporin**[‡] **Lorenzo Frattarolo**[§] **Loriana Pelizzon**[¶]

*We thank the participants at the International Symposium on Forecasting 2015, Riverside, USA; and the Computational and Financial Econometrics Conference 2015 in Pisa, Italy; the SYRTO Conference held in Amsterdam in 2015; the 79th International Atlantic Economic Conference held in Milan in 2015; the 9th Financial Risk International Forum of the Bachelier Institute held in Paris in 2016; the 3rd Annual Conference of the International Association for Applied Econometrics held in Milan in 2016; the SYRTO Conference held in Paris in 2016; the Econometrics and Statistics conference held in Hong Kong in 2017; the Systemic Risks in Financial Institutions held in Moscow in 2018; and seminar presentations at the Complutense University of Madrid and Tasmanian School of Business and Economics for the comments received from them. The usual disclaimers apply. The authors acknowledge financial support from the European Union, Seventh Framework Program FP7/2007-2013 under grant agreement SYRTO-SSH-2012-320270, the MIUR PRIN project MISURA - Multivariate Statistical Models for Risk Assessment, the SAFE Research Center, funded by the State of Hessen initiative for research, LOEWE, and the Department of Statistical Sciences of the University of Padova through the MIUR Project of Excellence "Statistical methods and models for complex data."

[†]University Ca' Foscari Venezia (Italy)

[‡]University of Padova (Italy) - Corresponding author, Department of Statistical Sciences, Via C. Battisti 241, 35121, Padova, massimiliano.caporin@unipd.it, +39-0498274199.

[§]University Ca' Foscari Venezia (Italy)

[¶]SAFE-Goethe University Frankfurt (Germany) and University Ca' Foscari Venezia (Italy)

Abstract

We propose a spatiotemporal approach for modeling risk spillovers using time-varying proximity matrices based on observable financial networks and introduce a new bilateral Multivariate GARCH specification. We study covariance stationarity and identification of the model, develop the quasi-maximum-likelihood estimator and analyze its consistency and asymptotic normality. We show how to isolate risk channels and we discuss how to compute target exposure able to reduce system variance. An empirical analysis on Euro-area bond data shows that Italy and Ireland are key players in spreading risk, France and Portugal are major risk receivers, and we uncover Spain's non-trivial role as risk middleman.

Keywords: spatial GARCH; network; risk spillover; financial spillover.

JEL Classification: C58, G10

1 Introduction

The US subprime and European sovereign bond crises sparked a renaissance in the research related to contagion and risk spillovers (Corsetti et al. (2011); Forbes (2012)) and it continues to be highly relevant. Within this research area, we aim to introduce an econometric model endowed with an economically grounded medium for variance and covariance spillovers[1]. We consider the simplest model for this medium by focusing on pairwise directed linkages summarized by a network. This allows us to introduce and exploit a parallel between the network approach and the tools commonly used in spatial econometrics.

Our proposal focuses on the risk dimension as Caporin and Paruolo (2015) with the advantage to extend the standard spatial econometrics definition of proximity. In fact, usually neighboring relations are time invariant, based on geographical measures of distance, as in Anselin (2013) and Elhorst (2003), or fixed economic properties, such as the industry sectors membership in Caporin and Paruolo (2015). Our network relationships are intended to be derived from specific and potentially granular financial variables (in

[1] see Bekaert and Harvey (1997); Ng (2000); Billio and Pelizzon (2003) for definitions of variance spillovers

our application, cross-border exposures of national banking systems). These suggested variables are those commonly perceived as potential time-varying transmission channels of shocks, and can thus characterize relationships in all their weighted, directional, and fluctuating generality. In particular, we take inspiration from a recent strand of spatial econometrics literature that includes finance-based weight matrices in the analysis (Keiler and Eder (2013), Blasques et al. (2016), Tonzer (2015), and Billio et al. (2017)). However, building on Caporin and Paruolo (2015), we differentiate from the latter papers since we focus on the risk dimension, as opposed to their aim to explain expected returns, and we include both contemporaneous (spatial) transmission effects and lagged cross-sectional (spatiotemporal) effects of the linkages, that are not considered in previous papers.

We propose an extension of the Spatial BEKK in Caporin and Paruolo (2015) named Spatial Bilateral BEKK (SB-BEKK) model. The main difference, as the name suggests, is the use of a bilateral specification for the proximity matrix, which introduces both Heterogeneity and Asymmetry in the model. In fact, the use of variable-specific proximity parameters allows heterogeneity in the network mediated effects. Heterogeneity was already present in the classical literature and Caporin and Paruolo (2015) revised it in a volatility framework. Furthermore, we have asymmetry of the adjacency (weight) matrix, which is typical of directed networks as considered in Billio et al. (2017); Tonzer (2015) and Blasques et al. (2016). Heterogeneity and Asymmetry were already used separately, but no one understood the consequences of their combination. Their joint consideration leads to different multiplication possibilities in proximity matrices, allowing the joint use of left and right multiplication parameters.

From a methodological point of view, the new bilateral specification requires a generalization of the identification condition stated in Caporin and Paruolo (2015). We are able to link the identification to the algebraic connectivity characteristics of an auxiliary undirected network, deriving a subtler condition than the one proposed in Caporin and Paruolo (2015)[2]. This sensitiveness of the identification condition to the connectivity

[2]Nevertheless, we show how in the symmetric case we recover their results

characteristics of the network sequence, implies that our bilateral specification is more responsive to actual network values. Indeed, thanks to our bilateral proximity, we are the first able to test with a simple linear restriction, the significance of each element of the dynamic BEKK matrices A and B, with a restricted BEKK specification.

A second distinctive feature of our framework is the proper treatment of the time dependence in the weight matrix. Previous applications of spatial methods to financial markets averaged these time-varying relationships, reducing everything to a static framework (see, as an example, Tonzer (2015)). We are the first to embed the time-varying nature of financial proximity within a nonlinear model for conditional covariance, going thus beyond the static weight matrix used in Caporin and Paruolo (2015). Recent advances in the linear modeling approach of spatial panels also address this point Lee and Yu (2012), Qu et al. (2017). Unfortunately, their assumption on the weight-matrix normalization is too restrictive in financial applications. Our methodology leads more appropriately to less constraining normalizations that add economic meaning to the methodology (see subsection 3.1.2). The essential ingredient for our time varying parameter recipe for a BEKK model is the joint spectral radius (JSR) of Rota and Strang (1960). We explain, with the aid of a simple Markov Switching example, how time variation in the matrix of parameters complicate the discussion and how the JSR represents a robust tool to obtain asymptotic properties. We derive covariance stationarity and QMLE consistency, paying the reasonable price of dealing with heterogeneous products of matrices, in place of matrix powers of the standard case, and computing bounds using the JSR. The proof of asymptotic normality greatly simplifies by the use of the same tools, but requires also our characterization of proximity in terms of spectral graph theory, see lemma 1.

With our model, we also take advantage of recent developments in spatial econometrics literature to summarize the behavior of the large number of series of covariance spillovers implied by model estimation. We also generalize the notion of direct and indirect effects to our bilateral specification and to the covariance framework LeSage and Pace (2014).

The last methodological improvement is an alternative, policy oriented, out of sample evaluation, based on counterfactual simulations. This procedure is less sensitive to the choice of covariance proxies needed in forecasting performance assessment. In addition, our alternative scheme represents a first step for policy evaluation of the optimal network configuration. A potential drawback of our proposal is that it can be criticized using the classical Lucas‚Äô argument. We discuss at length the issue and develop a parameter constancy test that could mitigate the effects of Lucas‚Äô critique.

We present an empirical analysis based on European sovereign CDS spreads considering a sample period that covers the collapse of Lehman Brothers, Greece's bailout, European sovereign crisis and Brexit. We focus on the role of cross-country banking system exposures in explaining the European sovereign CDS spillovers, and provide a counterfactual analysis evaluating the risk-reducing exposures on the period of the Brexit referendum.

Our empirical analysis leverages on our methodological innovation introducing a few elements: i) a normalization able to include rest of the world effects and to remove the effect of a deleveraging trend; a network map of significant exposure channels for spillovers; ii) an analysis of risk receivers and risk spreaders; iii) an evaluation of the diversification through the network of exposures; iv) a characterization of the exposure configuration that would have been able, according to our model, to reduce the variance during the Brexit referendum.

The paper is organized as follows. In Section 2, we introduce our new methodology detailing the econometric model, direct and indirect effects and the counterfactual analysis. In Section 3, we run an empirical analysis on CDS differences for the major Euro area countries during the subprime, sovereign debt crises, and Brexit. Finally, Section 4 summarizes our findings and concludes. The Appendix A includes definitions, corollaries, and proofs of the main results. The online supplementary material reports further methodological details, complementary tables and figures.

2 Spatial Econometrics of Network-Mediated Risk Spillovers

Three parts compose our framework for a spatiotemporal econometric treatment of risk and dependence relationships across entities. First, a multivariate GARCH model, in which lagged shocks propagate through fluctuating relationships. Second, a covariance decomposition able to investigate risk circulation and diversification in a synthetic way. Third, an evaluation of model relevance based on a simulated counterfactual experiment.

2.1 Spatial Bilateral BEKK

One of the benchmark models for conditional covariance is the BEKK GARCH of Engle and Kroner (1995). Unfortunately, even in its basic specification (the BEKK(1,1)), the model is computationally unfeasible, even for moderate values of n, due to its large number of parameters ($2n^2 + 0.5n\,(n+1)$). For this reason, the standard practice is to restrict the number of parameters, using either scalar or diagonal matrices. Despite being feasible, these common choices impose strong limitations on the detection of risk spillovers and variance feedbacks.

To overcome these critical aspects, Caporin and Paruolo (2015) introduce the Spatial-BEKK GARCH model. The spatial version of the BEKK model has the advantage of being more parsimonious than the full BEKK case, and, at the same time, it is more flexible than the diagonal specification, as it includes spillovers and feedback effects. We give a graph theoretic interpretation of their model that allow an extension including time-varying and asymmetric relationships. A weighted network (graph) is an ordered pair of sets and a function $G = (V, E, w)$ where $V = \{1, \ldots, n\}$ is the set of vertexes (or nodes), $E \subset V \times V$ the set of edges (or arcs), and $w(e) : E \mapsto \mathbb{R}^+$ is the weight function attributing strength to the edges. An edge between two nodes exists if there is a relationship between them and it can be identified as the (ordered) pair $\{i, j\}$ with $i, j \in V$. If there is no direction in the connection between nodes, then an edge $\{u, v\}$ is an unordered pair of nodes and the graph G is said to be undirected, whereas if a direction exists, then each edge $\{i,j\}$ is defined as an ordered pair of nodes and the graph G is said to be directed graph (or digraph). Different edges could have different strength as summarized by the weight function. The vertex adjacency structure of a n-order graph $G=(V,E,w)$ can be represented through a n-dimensional matrix W called an adjacency matrix. Each element $\omega_{i,j}$ of the adjacency matrix is equal to $w(\{i,j\})$ if there is an edge $\{i,j\}\in E$ (i.e., an edge from institution u to institution v with $i,j\in V$), and $\omega_{i,j}=0$ otherwise. If the graph is undirected, then $\omega_{i,j}=\omega_{i,j}$, that is, the adjacency matrix is symmetric. Common measures of node importance are in-degree , the number of incoming edges to the node $d^{in}_{G,i}=\sum_{i=1}^{n}\omega_{i,j}$ and out-degree, the number of outgoing edges from the node $d^{out}_{G,i}=\sum_{j=1}^{n}\omega_{i,j}$. In the undirected case in-degrees and out-degrees are the same and they are called degrees $d_{G,i}$.

Given a vector y_t of n cross-sectional observations at time t, we define $u_t=y_t-\bar{y}$, where $\bar{y}$ is the vector of sample means. Furthermore, we assume that a collection of time-varying adjacency matrices is available, i.e. all our analyses are conditional to an observed sequence of networks W_t, $t=1,2,\ldots T$. Our Spatial Bilateral BEKK GARCH (SB-BEKK) has the following structure:

$$u_t=\Sigma_t^{1/2}\epsilon_t \quad \epsilon_t\sim\mathcal{N}\left(0,I_n\right), \quad t=1,\ldots,T$$
$$\Sigma_t=CC'+A\left(W_t\right)u_{t-1}u_{t-1}'A\left(W_t\right)'+B\left(W_t\right)\Sigma_{t-1}B\left(W_t\right)'$$

where C is a lower triangular matrix, $\Sigma_t^{1/2}$ is the Cholesky decomposition of Σ_t,[3] and the parameter matrices have a specification described in the following equations:

$$\begin{aligned}A\left(W_t\right) &= A_0+A_{1,L}W_t+W_tA_{1,R}=\mathrm{diag}\left(a_0\right)I_n+\mathrm{diag}\left(a_{1,L}\right)W_t+W_t\mathrm{diag}\left(a_{1,R}\right)\\ B\left(W_t\right) &= B_0+B_{1,L}W_t+W_tB_{1,R}=\mathrm{diag}\left(b_0\right)I_n+\mathrm{diag}\left(b_{1,L}\right)W_t+W_t\mathrm{diag}\left(b_{1,R}\right),\end{aligned} \quad (1)$$

[3]Alternatively to the Cholesky, we can compute the square root by resorting to the spectral decomposition and set $\Sigma_t^{1/2}=D_tP_t^{1/2}D_t'$, where D_t is the matrix of eigenvectors and P_t is the diagonal matrix of eigenvalues.

where a_0, $a_{1,M}$, b_0 $b_{1,M}$, with $M = L, R$, are $n \times 1$ vectors.

In the spatial econometrics literature (see Anselin (2013), LeSage and Pace (2009), Elhorst (2003)), parameter matrices in the system (1) are called proximity matrices. In fact, in (1), the parameters are linearly dependent on the distances reported in the weight matrices W_t. In this literature, the prototypical example for W_t is a constant symmetric matrix detailing the real geographical distance.

Our model, instead, is able to handle time-varying asymmetric matrices and nests the vast majority of previously introduced specification of proximity matrices, including the heterogeneous impact specification, used in Caporin and Paruolo (2015), among the others. The right multiplying matrices $A_{1,R}$ and $B_{1,R}$ are a novelty introduced by our bilateral specification and exploit the asymmetry of adjacency matrix, proper of directed networks.

Extending the model of Caporin and Paruolo (2015) to time-varying asymmetric matrices, leads to a time-varying parameter model, making harder to derive stationarity of the process, and asymptotic properties of the estimator. In addition, the bilateral specification needs a different discussion of identification conditions, that depends on the fine details of the connectivity of the network sequence. In this respect, our network point of view cannot be seen as a simple redefinition of previous results, but allow us, to properly identify and estimate the model.

We also extend covariance stationarity, consistency and asymptotic normality in the case of time varying proximity using the Joint Spectral Radius. This is another innovation that could be useful in discussing asymptotic properties of different models.

We suggest to estimate the parameters of our SB-BEKK, conditional on the availability of the full sequence W_t for $t = 1, 2, \ldots, T$, by means of QMLE methods. If we denote by $\theta \equiv (\mathrm{vec}(C), a_{0,M}, a_{1,M}, b_{0,M}, b_{1,M})$ the vector of model parameters, the log-likelihood $L_T(\theta)$ is :

$$\ell_t(\theta) = \frac{n}{2}\log(2\pi) + \frac{1}{2}\det(\Sigma_t) + \frac{1}{2}u_t(\Sigma_t)^{-1}u_t' \quad , \quad L_T(\theta) = -\frac{1}{T}\textstyle\sum_{t=1}^{T}\ell_t(\theta) \ .$$

2.1.1 Network Connectivity in Identification

Unidentified models are characterized by a singular likelihood's Hessian, causing highly unstable QMLE numerical estimates. Identification restriction are essential to obtain a reliable inference based on QMLE. The identification condition of Caporin and Paruolo (2015) does not extend to our asymmetric model. This leads us to the, lead us to the discovery of an interesting relationship between parameters identification and network connectivity.

In the following proposition and lemma, we extend results included in Caporin and Paruolo (2015), with a detailed study of identification for the bilateral specification. Moreover, as the network matrices W_t might evolve on a time scale lower than that adopted for entities, as we will see in the empirical section, we allow for the presence of a collection of K distinct matrices W_k, $k = 1, 2, \ldots K$ with $K \leq T$.

Difficulties in identification of our model can be understood considering the example of a constant symmetric $W_k = W = W'$. In that case matrices commute

$$W \mathrm{diag}\left(a_{1,R}\right) = \sum_{i=1}^{n} a_{1,R,i}\omega_{i,j} = \sum_{j=1}^{n} a_{1,R,j}\omega_{i,j} = \mathrm{diag}\left(a_{1,R}\right) W' = \mathrm{diag}\left(a_{1,R}\right) W,$$

and our proximity matrix $A(W)$ is:

$$A(W) \;\; = \;\; \mathrm{diag}\left(a_0\right) I_n + \left(\mathrm{diag}\left(a_{1,L}\right) + \mathrm{diag}\left(a_{1,R}\right)\right) W.$$

The same proximity can be obtained using all the combination of left and right parameters that give equal element by element sum and the model is not identified. One possible solution is restricting right parameters to zero, reducing to Caporin and Paruolo (2015) original specification. Accordingly, our condition must be finer than the one proposed in this previous manuscript, and it must properly take into consideration the connectivity of a directed network sequence.

The study of identification is possible by making explicit the expression of $A(W_k)$ and

$B(W_k)$ as a linear function of a vector of parameters. Considering $A(W_k)$ as an example, we could write:

$$\operatorname{vec}(A(W_k)) = M(W_k)\begin{bmatrix} a_0 \\ a_{1,L} \\ a_{1,R} \end{bmatrix} \quad , \quad M(W_k) = \begin{bmatrix} I_n \otimes I_n & W_t' \otimes I_n & I_n \otimes W_k \end{bmatrix}(I_3 \otimes H)$$

where $H = \sum_{i=1}^{n} e_i \otimes e_i e_i'$ and e_i is the i-th column of I_n.

This explicit expression is novel with respect of Caporin and Paruolo (2015), even in the symmetric case. Moreover, an anonymous referee to whom we are particularly grateful pointed out how this linear map, in the asymmetric case, cannot be injective due to the rank deficiency of $M(W_k)$. A proper identification of the model is, then, impossible, without imposing further restrictions on the parameters.[4]

Before stating the identification conditions, we remind to the reader additional notions from network theory needed in the following. Given an undirected graph G the local structure and connectivity is encoded in the Laplacian matrix:

$$L(G) = \operatorname{diag}(d_G) - W \tag{2}$$

A connected component of G is a sub-graph in which all the nodes can be reached from any other node. A strongly connected component is a sub-graph in which all the nodes can be reached from any other node following edge direction. For undirected graphs all connected component are strongly connected. An isolated node is considered a strongly connected component.

Starting from each G_k define an auxiliary network sequence $\mathfrak{G}_k = (V, \mathfrak{E}_k, \mathfrak{w}_k)$, on the same nodes, but with undirected edges and different weights, according to the adjacency

[4] We reproduce conditions in Caporin and Paruolo (2015) for the symmetric case; see corollary 1 in Appendix subsection A.1

matrix $\mathfrak{W}_k$:

$$\mathfrak{W}_k = \left(W_k' \odot W_k'\right)\left(D_{W_k \odot W_k}\right)^{-1}\left(W_k \odot W_k\right) \quad , \quad D_{W_k \odot W_k} = \operatorname{diag}\left(\left(W_k \odot W_k\right)\mathbf{1}\right).$$

We define as c_k the number of connected components in $\mathfrak{G}_k$ of which s_k isolated nodes without self-loops.

The next lemma is the core result linking the proximity matrix to network connectivity. It is needed for identification and asymptotic normality.

Lemma 1 (Spectrum of $M'M$)**.** *$M\left(W_k\right)' M\left(W_k\right)$ has n eigenvalues equal to 1 , n eigenvalues equal to the degrees $d_{\mathfrak{G}_k,i}$, n eigenvalues equal to the eigenvalues of $L\left(\mathfrak{G}_k\right)$. Suppose that $\mathfrak{G}_k$ has c_k connected components of which s_k are isolated nodes without self loops. Then $M\left(W_k\right)' M\left(W_k\right)$ has $c_k + s_k$ null eigenvalues.*

Next we propose our identification Theorem:

Theorem 1 (Identification)**.** *Assume that at least one of the matrices W_k is not symmetric. Let π_K correspond to the vectorized collection of either $A\left(W_k\right)$ or $B\left(W_k\right)$ for $k = 1, 2, \ldots K$, that is $\pi_K = vec\left(\left[A\left(W_1\right)' : A\left(W_2\right)' : \ldots : A\left(W_K\right)'\right]'\right)$, where $A\left(W_k\right)$ or $B\left(W_k\right)$ matrices are those coming from a full BEKK model fitted on each sub-sample, and are thus globally identified. Let ψ be the parameter vector in the SB-BEKK representation corresponding to π_K. Additionally, let $\chi = \min_{1,\ldots,K} c_k + s_k$. Then, a necessary and sufficient condition for the identification of ψ is that we place χ linear restriction on it.*

In the next lemma a sufficient condition for a single restriction is given on the original network sequence G_k.

Lemma 2 (Single Restriction)**.** *A sufficient condition to have $\chi = 1$ is that at least one of the original weight matrices W_k is fully indecomposable, inducing an ultrastrong graph in the sense of Brualdi (1967).*

The condition of Lemma 2 is not restrictive for small complete networks such as the one we use in our empirical analysis. For bigger and sparser networks sequences, the number of components of the induced networks sequence $\mathfrak{G}_k$ could be checked using, for example, Tarjan's algorithm (Tarjan, 1972). In figure 1, we give a simple example, where G has single strongly connected component but the induced $\mathfrak{G}$ has three.

[Insert figure 1 around here]

In the full BEKK(1,1), it is sufficient to set $A_{1,1}, B_{1,1} > 0$ to have global identification, and in our case, this is equivalent to the conditions $a_{0,1} > 0$ and $b_{0,1} > 0$. Then, under Lemma 2, we choose to achieve identification by imposing an equal sum of left and right parameters:

$$a_{1,R,n} = -\sum_{i=1}^{n} a_{1,L,i} + \sum_{i=1}^{n-1} a_{1,R,i} \quad , \quad b_{1,R,n} = -\sum_{i=1}^{n} b_{1,L,i} + \sum_{i=1}^{n-1} b_{1,R,i}$$

We stress that this identification strategy allows for the presence of coefficients of both signs in $a_0, b_0, a_{1,L}, a_{1,R}, b_{1,L}$ and $b_{1,R}$. In subsection 2.2, we show how this allows contributions that reduce the variance. Subsection A.1 in the Appendix reports the proof of the previous theorem and lemmas, alongside corollary 1 for symmetric W_k matrices.

2.1.2 Joint Spectral Radius in Covariance Stationarity

A necessary and sufficient condition for stationarity and geometric ergodicity, and thus a sufficient condition to ensure the ergodicity and strict stationarity of the process implied by the model, is covariance stationarity, Boussama et al. (2011). According to the discussion in Avarucci et al. (2013), this condition, distinct from the univariate case, appears necessary for consistency and asymptotic normality of the QMLE estimator. To show the covariance stationarity, we introduce a VARMA representation of the process. In addition, the condition for covariance stationarity will be expressed using the joint spectral radius (JSR) of Rota and Strang (1960). It is well known that a BEKK model admits a VARMA representation (see, for example, Hafner and Rombouts (2007)). In our case, the VARMA coefficients are time varying because they depend on the network sequence, and ensuring covariance stationarity requires the study of convergence of a geometric series with heterogeneous terms, analogous to the one studied in the generalized autoregressive model; see Brandt (1986) and Bougerol and Picard (1992).

Heterogeneity requires a specific handling of stationarity as we discuss in the following instructive example.[5] We introduce a 2 state Markov Switching bivariate VAR(1) model. The series of coefficient matrices follows a Markov chain, showing a non-trivial autocorrelation pattern as we expect from coefficients deriving from a series of financial networks. Conditionally on the realized sequence of networks the process is a VAR. The conditional process is then similar to the conditional VARMA representation that we will give of the SB-BEKK. In addition, sharp sufficient conditions for covariance stationarity of this model can be found in Francq and Zakoian (2002). The model is the following:

$$X_t = \mu + \Phi_{s_t} X_{t-1} + \epsilon_t, \quad \mu = \frac{1}{2}\mathbf{1}_2, \quad \epsilon_t \sim \mathcal{N}\left(\mathbf{0}_2, I_2\right)$$

Let us define six matrices

$$\Psi_0 = \frac{2}{3}\begin{pmatrix} \cos\left(\frac{3}{2}\right) & \sin\left(\frac{3}{2}\right) \\ -2\sin\left(\frac{3}{2}\right) & 2\cos\left(\frac{3}{2}\right) \end{pmatrix} \quad \Psi_1 = \frac{2}{3}\begin{pmatrix} 2\cos\left(\frac{3}{2}\right) & 2\sin\left(\frac{3}{2}\right) \\ -\sin\left(\frac{3}{2}\right) & \cos\left(\frac{3}{2}\right) \end{pmatrix}$$

$$\Xi_0 = \frac{2}{3}\begin{pmatrix} \cos\left(\frac{3}{2}\right) & \sin\left(\frac{3}{2}\right) \\ -2\sin\left(\frac{3}{2}\right) & 2\cos\left(\frac{3}{2}\right) \end{pmatrix} \quad \Xi_1 = \frac{2}{3}\begin{pmatrix} 2\cos\left(\frac{3}{2}\right) & -2\sin\left(\frac{3}{2}\right) \\ \sin\left(\frac{3}{2}\right) & \cos\left(\frac{3}{2}\right) \end{pmatrix}$$

$$P = \begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix} \quad Q = \begin{pmatrix} 0.6 & 0.4 \\ 0.4 & 0.6 \end{pmatrix}$$

and consider the four specifications:

1. $\Phi_{s_t} = \Psi_{s_t}$, $\mathbb{P}\left(s_t \middle| s_{t-1}\right) = P$
2. $\Phi_{s_t} = \Psi_{s_t}$, $\mathbb{P}\left(s_t \middle| s_{t-1}\right) = Q$
3. $\Phi_{s_t} = \Xi_{s_t}$, $\mathbb{P}\left(s_t \middle| s_{t-1}\right) = P$
4. $\Phi_{s_t} = \Xi_{s_t}$, $\mathbb{P}\left(s_t \middle| s_{t-1}\right) = Q$

[5] Adapted from Jungers (2009) to a stochastic framework.

It is well known that a VAR(1) is covariance stationary if the maximum absolute value of the eigenvalues (spectral radius $\rho(\Phi)$) of its coefficient matrix Φ is less than one. We have $\rho(\Phi_0) = \rho(\Phi_1) = \rho(\Xi_0) = \rho(\Xi_1) = 0.9428$ and, naively, one can infer that the switching model inherit covariance stationarity, in all specifications.

[Insert figure 2 around here]

Figure 2 shows simulated paths of the two variables of the model in case the states are fixed or switching. The series in the bottom left panel, corresponding to the switching specification $\Phi_{s_t} = \Psi_{s_t}, \mathbb{P}(s_t | s_{t-1}) = Q$, clearly exhibit non stationarity. This counter intuitive result has the following explanation. The spectral radius of the product $\rho(\Psi_0\Psi_1) = \rho(\Psi_1\Psi_0) = 1.7510$, is bigger than one. The auto-covariance function of the model conditional on the realized chain will be a function of the product of the realized matrix coefficients. The unconditional stationarity of the model is going to depend on the ratio of probability of occurrence in each possible matrix product sequence of the squares Ψ_0^2, Ψ_1^2 that favor stationarity , and of the mixed products $\Psi_0\Psi_1, \Psi_1\Psi_0$ that favor non stationarity. Given the transition matrix P of the chain, alternating occurrence is more probable, in this particular specification, and the non stationary character prevail. This example shows that in the heterogeneous case is important to take into consideration the spectral characteristic of all the possible products of the realization of matrix coefficients. The JSR was designed with this scope. In particular, it is possible to show that the JSR is an upper bound for the spectral radius of all possible products.

To introduce it, let us consider an infinite countable set of $n \times n$ matrices $\mathcal{A} = \{A_i\}_{i=0}^{\infty}$ with the convention that $A_0 = I_n$. A generic product of t elements from $\mathcal{A}$ could be obtained by extracting uniformly, with replacement, t elements from $\mathcal{A}$ and matrix-multiplying them. For example, suppose that the elements sampled have indexes $\sigma_1 = 44, \sigma_2 = 44, \sigma_3 = 20, \ldots, \sigma_t = 1$; the product will be $A_{44}^2 A_{20} \cdots A_1$. Let us define the set of all those possible products $\mathcal{A}^t = \left\{M \in \mathcal{A}^t | M = \prod_{i=1}^{t} A_{\sigma_i}, s.t. A_{\sigma_1} \in \mathcal{A}, \ldots A_{\sigma_t} \in \mathcal{A}\right\}$. We have:

Definition 2.1. *Joint Spectral Radius. Given a proper norm $\|\cdot\|$ on $\mathbb{R}^n \times \mathbb{R}^n$, we define* *on the set $\mathcal{A}$ the joint spectral radius $\varrho(\mathcal{A})$ by*

$$\varrho(\mathcal{A}) = \lim_{t\to\infty} \hat{\rho}_t(\mathcal{A}) \tag{3}$$

$$\hat{\rho}_t(\mathcal{A}) = \sup_{A_{\sigma_1},\dots,A_{\sigma_t}\in\mathcal{A}} \left(\left\|\prod_{i=1}^{t} A_{\sigma_i}\right\|\right)^{1/t} \tag{4}$$

The JSR depends only on the set of matrices included in the sequence with no reference to the probability measure associated to it. Even if a bound on the spectral radius is a strong requirement, it is robust with respect to the choice of probability measure. For our conditional modeling this is an essential feature as we do not detail the underlying data generating process (DGP) of the network sequence.

Let us clarify the previous statement referring to our Markov Switching example. Considering Ξ_0 and Ξ_1 the spectral radius of the mixed product is less than one ($= 0.9345$), the JSR of the set $\{A_0, A_1\}$ is also bounded by one. Independently from the probability measure i.e. from the choice of P or Q as a transition matrix, the process would be stationary in covariance. This can be seen in the bottom right part of figure 2. Instead, in the cases, with $\Phi_{s_t} = \Psi_{s_t}$, covariance stationarity depends on the choice of measure induced by the different transition matrices: stationary with P, non stationary with Q (see bottom left part of figure 2).[6] So, an unitary bound on the JSR can be more restrictive than needed, but it is robust against any possible DGP of the random matrix sequence that satisfies it. Since in the following we condition on the matrix sequence without specifying its DGP, bounds on the JSR represent a suitable approach.

We are now ready to introduce the VARMA representation of the SB-BEKK(1,1):

$$X_t = \tilde{C} + \left(\tilde{A}(W_t) + \tilde{B}(W_t)\right) X_{t-1} - \tilde{B}(W_t)\,\eta_{t-1} + \eta_t \tag{5}$$

[6] Covariance stationarity of the example could be checked using the sharper bounds in Francq and Zakoian (2002)

where

$$
\begin{aligned}
X_t &= \operatorname{vech}\left(u_t u_t'\right), \quad \xi_t = \operatorname{vech}\left(\Sigma_t\right), \quad \eta_t = X_t - \xi_t, \quad \tilde{C} = \operatorname{vech}\left(CC'\right) \\
\tilde{A}\left(W_t\right) &= L_n\left(A\left(W_t\right) \otimes A\left(W_t\right)\right) D_n, \quad \tilde{B}\left(W_t\right) = L_n\left(B\left(W_t\right) \otimes B\left(W_t\right)\right) D_n,
\end{aligned}
$$

and L_n is the elimination matrix, while D_n is the duplication matrix (see Magnus and Neudecker (1999)). Given the distribution of ϵ_t in (1), adopting the terminology of Hafner and Rombouts (2007), we have that u_t is a strong GARCH process and η_t is a martingale difference sequence. This is true even if the innovation distribution is misspecified, but remains i.i.d. Equation (5) puts the model outside of the general specifications of dynamic spatial panel models given in Elhorst (2001), because of the presence of a spatiotemporal moving average term.

Given the discussion at the beginning of the subsection the following theorem is not a surprise:

Theorem 2 (Covariance Stationarity)**.** *If the joint spectral radius*

$$
\varrho\left(\left\{\tilde{A}\left(W_t\right) + \tilde{B}\left(W_t\right)\right\}_{t=-\infty}^{\infty}\right) < 1,
$$

the SB-BEKK(1,1) process is covariance stationary.

We stress that the use of bounds on the joint spectral radius is a condition on dynamic stability strictly weaker than the ones already present in the spatial econometrics literature. For example, the uniform boundedness assumption size in Lee and Yu (2012), being based on row and column sum norm, implies our JSR condition. This is relevant for our empirical application where the choice of normalization is different from the standard row sum, and motivated by the economics of the data used (see subsection 3.1.2).

2.1.3 Inference and Networks

The possibility of inference on single elements of time varying coefficient matrices with a single linear restriction is one of the main features of our left-right specification not accessible to any previous spatial model in the literature. Using a simple linear restriction, we can test whether an off-diagonal element of $A(W_t)$ (or $B(W_t)$) is statistically significant:

$$H_{0,i,j} : [A(W_t)]_{i,j} = (a_{1,L,i} + a_{1,R,j})\, \omega_{t,i,j} = 0 \qquad i \neq j \tag{6}$$

Then, if we are able to show consistency and asymptotic normality of the QMLE estimator, conditionally on the network sequence, the test statistic for $H_{0,i,j}$ is asymptotically distributed as a standard normal. In our empirical application, we propose a network representation of results from the test, picturing the significance and relevance of different channels of risk circulation. We now discuss the consistency and asymptotic normality of the QMLE estimator that make this mapping of risk flows possible.

We follow closely Hafner and Preminger (2009) proof and substitute products of matrices in place of simpler powers of matrices, whenever it is the case. Accordingly, the bounds needed to obtain the results are expressed in terms of joint spectral radii, making the conditional results valid for all the DGPs of the network sequence that respect our covariance stationarity condition. In addition, as already remarked, in the proof of asymptotic normality, bounds on likelihood derivatives depend on bounds on the spectral radius of $M(W_k)' M(W_k)$ requiring again the network perspective of lemma 1.

In the following, $\|\cdot\|$ represents the norm operator, with different norms being specified when needed. Denote by θ the vector of stacked parameters that implicitly satisfy the identification condition of Theorem 1, and denote the true parameter vector as θ_0. Define the QMLE as $\theta_T = \arg\max_{\theta \in \theta} L_T(\theta)$. Additionally, let $\tilde{\Sigma}_t$ be the process where the starting values are drawn from their stationary distribution, and let $\tilde{\xi}_t$, $\tilde{L}_T$ and $\tilde{\ell}_t$ be defined analogously.

We begin by discussing the assumptions needed for consistency.

Assumption 1. *The parameter space* Θ *is compact and* $\varrho\left(\left\{\tilde{A}\left(W_t\right)+\tilde{B}\left(W_t\right)\right\}_{t=-\infty}^{\infty}\right)<1$.

Assumption 2. $\{u_t\}$ *is strictly stationary and ergodic, and* $\exists s>0$ *s. t.* $\mathbb{E}\left[\|u_t\|^s\right]<\infty$.

Assumption 3. $\mathbb{E}\left[\|\epsilon_t\|^s\right]<\infty,\ Var\left[\epsilon_t\right]=I_n$.

Assumption 4. *The model is identified: i.e., conditions in Theorem 1 are satisfied.*

Our assumptions parallel those in Hafner and Preminger (2009). The only conceptual difference, once we take into account our use of time-varying matrices, is in Assumption 1, where we bound the JSR of the sum of $\tilde{A}\left(W_t\right)$ and $\tilde{B}\left(W_t\right)$, instead of a condition that points only at the JSR of $\tilde{B}\left(W_t\right)$ matrices. In Lemma 4 in subsection A.2 of the Appendix, we show that one condition implies the other. This was done for the standard BEKK in Boussama et al. (2011), where they also show that bounding the spectral radius of the sum was a sufficient condition for strictly stationarity and ergodicity. Then, in the standard framework, the first part of assumption 2 would be redundant. Showing that this also applies to our case is outside the scope of the paper.

Theorem 3 (Consistency)**.** *Under Assumptions 1-4* $\hat{\theta}_T\rightarrow_{a.s.}\theta_0$.

To establish asymptotic normality, the following additional assumptions are needed.

Assumption 5. *The parameter* θ_0 *is an interior point of* Θ

Assumption 6. $\mathbb{E}\left[\|u_t\|^6\right]<\infty$

Assumption 7. $\sup\max_{i=1,\ldots,n}\sum_{j=1}^{n}W_{t,ij}\leq d^*<\infty$ *a.s.*

Assumptions 5 and 6 are identical to assumptions adopted in Hafner and Preminger (2009). Assumption 7, pointing at the network structure, is not particularly restrictive. In fact, it is trivially verified for the row-normalization case, but it also justified for different normalization schemes. For example, the network used in our empirical analysis has $d^*=1$ (c.f. section 3.1.2).

Further, let us also define the following matrices

$$V=\mathbb{E}\left[\frac{\partial\tilde{\ell}\left(\theta_0\right)}{\partial\theta}\frac{\partial\tilde{\ell}\left(\theta_0\right)}{\partial\theta'}\right],\qquad J=\mathbb{E}\left[\frac{\partial^2\tilde{\ell}\left(\theta_0\right)}{\partial\theta\partial\theta'}\right].$$

Theorem 4 (Asymptotic Normality)**.** *Under Assumptions 1-7*

$$\sqrt{T}\left(\hat{\theta}_T - \theta_0\right) \rightarrow_D N\left(0, J^{-1}VJ^{-1}\right).$$

2.2 Direct, Indirect and Mixed Spillovers

The introduction of proximity matrices in the dynamic of BEKK models allows the estimation $0.5(n+1)n$ series of filtered conditional covariance elements. For $n \geq 3$, it is difficult to interpret directly all the recovered series, and it is therefore desirable to have summary measures backed by some theoretical line of reasoning. This is a classical issue in spatial econometrics, where we observe the same difficulty in interpreting the impact of explanatory variables or innovations. The complexity stems from the large cross-sectional dimension of the analyzed data (or series), as in our case. The traditional solution is to resort to summary measures of the direct and indirect effects of explanatory variables and shocks; see LeSage and Pace (2009) and LeSage and Pace (2014). The SB-BEKK framework has an additional layer of complexity. As we discuss in the following, left and right multiplication allow researchers to focus on different aspects of risk propagation. To better understand this model attitude, it is advisable to recall the notions of direct and indirect effects of shock diffusion, previously introduced in the spatial econometrics literature (see LeSage and Pace (2014)) and generalized here for the SB-BEKK model. The starting point is the Spatial Error Model (SEM), where the $n-$variate dependent variable v_t depends on an $n-$dimensional vector of shocks u_t, on a time invariant (for simplicity) weight matrix W, and on a scalar parameter θ

$$v_t = \left(I_n + \theta W\right) u_t.$$

LeSage and Pace (2014) decompose the error term in the direct effect v_t^0 and the local indirect effect v_t^1 as follows:

$$v_t = v_t^0 + v_t^1 \quad , \quad v_{i,t}^0 = [I_n u_t]_i = u_{i,t} \quad , \quad v_{i,t}^1 = [\theta W u_t]_i = [W \theta u_t]_i = \theta \sum_{j=1}^{n} \omega_{i,j} u_{j,t}$$

where $[X]_{i,j}$ identifies the element of position i, j of the argument matrix X with one single index if X is a vector, $\omega_{i,j}$ represents the "distance"between subject i and subject j coming from the spatial weight matrix W, and by definition $\omega_{i,i} = 0$.

This means that the target variable $v_{i,t}$ depends on its own shock, as monitored by v_t^0, the direct impact. Further, it is also affected by the indirect impact v_t^1. The latter captures the effect coming from neighboring elements $v_{j,t}$ with $i \neq j$ and with an impact only from those j such that $\omega_{i,j} \neq 0$. We note that in the SEM model, left and right multiplication are identical due to the presence of a scalar parameter θ. We translate these elements into the SB-BEKK model and provide a novel decomposition. We focus on the ARCH part of the model because we want to highlight the role of innovations. We note that

$$\begin{aligned} v_t = A(W) u_t &= (A_0 + A_{1,L} W + W A_{1,R}) u_t = v_t^0 + v_{L,t}^1 + v_{R,t}^1 && (7) \\ v_{i,t}^0 &= [A_0 u_t]_i = a_{0,i} u_{i,t} && (8) \\ v_{L,i,t}^1 &= [A_{1,L} W u_t]_i = a_{1,L,i} \sum_{j=1}^{n} \omega_{i,j} u_{j,t}. && (9) \\ v_{R,i,t}^1 &= [W A_{1,R} u_t]_i = \sum_{j=1}^{n} \omega_{i,j} a_{1,R,j} u_{j,t}. && (10) \end{aligned}$$

The $i-$th element of v_t depends on its own past shock, weighted by the coefficient $a_{0,j}$ (direct effect), on the past shocks of its neighbors weighted by the distance, loaded with the sum of the same coefficient, $a_{1,L,j}$ (indirect left effect), and a coefficient different from each source $a_{1,R,i}$ (indirect right effect). For the GARCH part, similarly to the ARCH case, we introduce:

$$m^{0}_{i,t} = [B_0 u_t]_i = b_{0,i} u_{i,t} \tag{11}$$

$$m^{1}_{L,i,t} = [B_{1,L} W u_t]_i = b_{1,L,i} \sum_{j=1}^{n} \omega_{i,j} u_{j,t} \tag{12}$$

$$m^{1}_{R,i,t} = [W B_{1,R} u_t]_i = \sum_{j=1}^{n} \omega_{i,j} b_{1,R,j} u_{j,t}. \tag{13}$$

Consequently, bearing in mind that we are discussing properties of a conditional covariance model, the left multiplication term allows us to investigate which are the risk receivers. In the right multiplication term, distinct from the left multiplication case, the coefficients in the indirect effect are not pointing at the subject we are monitoring (subject j) but at the subject originating the shock (subject i). With the right multiplication term, the parameters magnify the effect of the sources of risk, allowing us to focus on risk spreaders.

Given the distinction between spreader in receivers possible in our model we need a suitable refinement of decomposition introduced in the spatial econometrics models

In addition, focusing on conditional covariance decomposition matrices, we have a breakdown conditional to the past, while usually it considers only contemporaneous variables, and we deal with quadratic forms.

Our decomposition will then be built of quadratic blocks depending on lagged variables. As an example we detail the contribution mediated by the interaction of two indirect left effect in the ARCH and GARCH part:

$$v^{1}_{L,i,t-1}, v^{1}_{L,j,t-1} = \left[A_{1,L} W \Sigma_{t-1} W' A'_{1,L}\right]_{i,j} \tag{14}$$

$$\left[\Omega^{1,1}_{L,L,t-1}\right]_{i,j} = \mathbb{C}\mathrm{ov}\left(m^{1}_{L,i,t-1}, m^{1}_{L,j,t-1} \middle| I_{t-2}, W\right) = \left[B_{1,L} W \Sigma_{t-1} W' B'_{1,L}\right]_{i,j}, \tag{15}$$

where I_{t-2} is the information set till $t-2$.

We propose a four-term decomposition of the system-conditional covariance:

1. *Constant Contribution*: This represents the part of the covariance that is unrelated

to the model dynamic and independent from both the network and the time;

2. *Direct Contribution*: This represents the covariance contribution from each entity's own past; it is the variance due to past direct effects, and therefore has no dependence on the network;

3. *Indirect contribution*: This represents the covariance contribution due to indirect effects that are due , only, to the assets' network exposures;

4. *Mixed contribution*: This represents the covariance contribution originating from the quadratic form of the model, and is due to the interaction of both direct and indirect elements.

[Insert table 1 around here]

In Table (1), we summarize the elements appearing in the conditional covariance decomposition.

We highlight that the decomposition is time varying by construction and might also be affected by the dynamic in the network structure. Since the model specifications allow for positive and negative signs on both ARCH and GARCH coefficients, in principle, diversification benefits could arise from all contributions less the direct one. The variance decomposition outlined above are specific to a single element of the covariance matrix. Of particular interest in the empirical section will be the decomposition of the variance of single series, a breakdown able to highlight which entity risk is the most affected by the presence of the exposure network (i.e., which are the more fragile nodes in the network). However, we might be interested in recovering a synthetic measure of the decomposition at the entire covariance level. We propose to define this synthetic (and time-varying) measure starting from a portfolio representation of the system, with portfolio weights given by the vector $\mathbf{z}$, such that $\sum_{i=i}^{n} z_i = 1$, and therefore leading to the following variance contributions (VC):

$$\begin{aligned}\sigma_t^2 &= \mathbb{V}\mathrm{ar}\left(\mathbf{z}' y_t \middle| \mathcal{I}_{t-1}\right) \\ &= VC_t^{Constant} + VC_t^{Direct} + VC_t^{Indirect} + VC_t^{Mixed}. \end{aligned} \tag{16}$$

The previous decomposition is able to report, globally, the importance of the network sequence through time. Another possibility is instead to compute the marginal spillover contributions $MSC_{k,t}$ that we could attribute to entity k:

$$\begin{aligned}MSC_{k,t} &= \frac{1}{2}\frac{\partial}{\partial z_k}\mathbb{V}\mathrm{ar}\left(\mathbf{z}' y_t \middle| \mathcal{I}_{t-1}\right) \\ &= MSC_{k,t}^{Constant} + MSC_{k,t}^{Direct} + MSC_{k,t}^{Indirect} + MSC_{k,t}^{Mixed} \end{aligned} \tag{17}$$

The marginal spillover contribution is normalized, guaranteeing that, if we define the vector $\mathbf{MSC}_t' = [MSC_{1,t}, \ldots, MSC_{n,t}]'$, we have $\mathbf{z}'\mathbf{MSC}_t = \mathbb{V}\mathrm{ar}\left(\mathbf{z}' y_t \middle| \mathcal{I}_{t-1}\right)$. $MSC_{k,t}$ is, then, a measure of the importance of node k as a source of spillovers in time t. In particular, our decomposition allows us to disentangle the direct sources of spillovers from the ones that are mediated by the network sequence.

Among the many possible choices, in the empirical application, we chose the simplest one, and thus consider a portfolio characterized by equal weights for each entity. This is also equivalent to considering the behavior of an average element of the covariance (i.e., the average linear dependence in the system). The use of different weighting schemes, with potentially better economic explanations, is left for further empirical research. The explicit expressions for variance decomposition and marginal spillover contributions are reported in the supplementary material.

2.3 Counterfactual Network and Covariance Reduction

Evaluating the out of sample accuracy of multivariate volatility models is a difficult task. One of the main issue is given by unobservable nature of the target and the need of a volatility proxy. In particular volatility proxies are usually computed on intra-day data and according to the literature Laurent et al. (2012) the higher the frequency the higher the possibility to discriminate among the models. Unavailability of high frequency data then requires alternative methodologies, less dependent on the choice of volatility proxy. In this paper we propose an approach tailored to our model, that has also the relevant feature of being policy oriented. In particular, we rely on a simulated counterfactual analysis (Mccallum (1988), Rotemberg and Woodford (1997)) to investigate how changes in the network, impact the variance of an equally weighted index. We change the network in order to minimize the variance forecast based on ex-ante data. This choice of the objective function is motivated by its equivalence with the minimization of the average forecasted covariance coefficent (i.e., on the average linear dependence and risk in the system). The counterfactual innovation paths are bootstrapped from estimated innovations (see section S of the Supplementary materials for the details of the circular bootstrap Politis and Romano (1992) we used). In a similar way, we recover a realized innovation path on the out-of-sample period using the previously estimated parameters and the out-of-sample observed network.

In the following, we describe the details of the procedure. Conditional on the bootstrapped innovations $\tilde{\epsilon}^{[b]}_{T+l}$ with $b \in [1, \ldots, N_B]$ and $l \in [1, \ldots, h]$, and assuming that the network is constant over the forecast horizon, the forecasted covariance path is a function of the network at time T, $W_T \mapsto \hat{\Sigma}^F_{T+l}(W_T)$ $l \in [1, \ldots, h]$. This raises the interesting possibility of obtaining a target network that can reduce the future risk and dependence in the system. To define the optimal target network, we require that it minimizes, at least locally, the average covariance coefficient of the system, which we compute as the variance of an equally weighted portfolio of all the series. We underline here that the methodology can easily accommodate different objective functions, being based on numerical optimization. A detailed investigation of more refined and economically motivated targets is left for future research. With our approach, we exploit the frequency mismatch between the data used to estimate the network and the series for which the risk is evaluated. Such situations are not rare, as financial networks might be built from lower-frequency data (using, for instance, balance-sheet data), while financial market data are available at a daily or even higher frequency. In particular, we assume that the network changes every q observations. That is, in the full sample T, we have $[T/q] = Q$ networks, or, alternatively, we have Q sub-periods in which the network is stable. In the forecast exercise, we assume that $W_{T+l} = W_Q$ for each $l \in [1, \ldots, h]$, such that $T+1$ and $T+h$ are the beginning and end of the period $Q+1$. We require that the average forecasted variance of the equally weighted index over period $Q+1$; i.e., the first sub-period following the estimation sample, conditional on the bootstrapped innovations, is minimized by numerically solving the following constrained optimization problem:

$$\min_{\text{vec}W^\star} \left\{ \frac{1}{h} \sum_{l=1}^{h} \frac{1}{n^2} \mathbf{1}' \hat{\Sigma}^F_{T+l} \left(W^\star \right) \mathbf{1} \right\}$$
$$s.t. \quad 0 \leq \left[W^\star \right]_{i,j} \leq 1 \quad for \quad i,j = 1 \ldots n, \quad \text{and} \quad \text{Tr} \left(W^\star \right) = 0$$

where $\mathbf{1}$ is the $n \times 1$ column vector whose elements are all equal to 1 and $\text{Tr}(.)$ is the trace operator. It is important to note that the estimated network $W^\star$ is weighted and directed, but is totally unrelated to the last available network. We thus also consider a more realistic constraint in which the out (in) strengths of the nodes, defined as the row (column) sums of the optimal network, are set to be the same as the out (in) strengths of the nodes of the last network W_Q. For the row-sum case, we impose

$$d^{out}_{W^\star,i} = d^{out}_{W_Q,i}, \tag{18}$$

and we can write a similar constraint for the column sum. These constraints avoid a change in the strengths of the nodes and correspond to a simple redistribution of the weights across the system. We refer to the constraint in equation (18) as a redistribution constraint. Moreover, we introduce an additional alternative constraint imposing that the resulting optimal network differs from the previous one in an ordinary way, and does not represent an exceptional change. We implement this by computing the Frobenius norm on the historical changes in the network

$$\left\|\Delta W_q\right\|_F = \left\|W_q - W_{q-1}\right\|_F = \sqrt{\operatorname{vec}\left(W_q - W_{q-1}\right)' \operatorname{vec}\left(W_q - W_{q-1}\right)} \qquad q = 1, \ldots Q-1, \tag{19}$$

and imposing that the norm of optimal change $\left\|\Delta W^\star\right\|_F = \left\|W^\star - W_Q\right\|_F$ is less than or equal to the empirical 0.95 quantile of historical norm changes $q_{0.95}^{\|\Delta W\|_F}$

$$\left\|\Delta W^\star\right\|_F \leq q_{0.95}^{\|\Delta W\|_F}. \tag{20}$$

We refer to the constraint in equation (20) as a Frobenius norm constraint.

To evaluate the performance of the proposed out-of-sample methodology, we suggest comparing two estimates of the model, one excluding the out-of-sample data, and the second including the forecasted data. This enables us to compute the filtered innovations for the forecasted periods, conditional on the true, observed $Q+1$ network:

$$\hat{\epsilon}_{T+l} = \hat{\Sigma}_{T+l}^{-\frac{1}{2}} \left(W_{Q+1}\right) u_{T+l} \qquad l \in [1, \ldots, h]. \tag{21}$$

Then we can reconstruct the us and the counterfactual proxy for the equally weighted index's conditional variance, as if the realized network for the period of interest is the optimal one, $W^\star$:

$$\begin{aligned} \tilde{u}_{T+l}^\star &= \hat{\Sigma}_{T+l}^{\frac{1}{2}} \left(W^\star\right) \hat{\epsilon}_{T+l} & (22) \\ \mathbb{V}\mathrm{ar}\left(\frac{1}{n}\mathbf{1}' y_{T+l}^\star \middle| \mathcal{I}_{T+l-1}\right) &= \mathbb{V}\mathrm{ar}\left(\frac{1}{n}\mathbf{1}' u_{T+l}^\star \middle| \mathcal{I}_{T+l-1}\right) \simeq \left(\frac{1}{n}\mathbf{1}' u_{T+l}^\star\right)^2 & (23) \end{aligned}$$

In this way, we can compare the obtained optimal counterfactual proxy with the realized proxy, the latter being robust against model misspecification. in addition, relative comparison of the same type of proxy in the two cases, make the procedure less dependent on its choice. If we observe a reduction of the optimal counterfactual proxy with respect to the realized one, we can conclude that, indeed, we can reduce the variance by optimally changing the network based only on ex-ante data. Even if this simulated counterfactual analysis is similar to the one used in monetary-policy evaluation exercises (see Rotemberg and Woodford (1997), Primiceri (2005) and more recently, Bikbov and Chernov (2013)), we are aware that, concerning our empirical application, regulators have, at the moment, no possibility of intervention in incentivizing the redistribution of banking exposures. So our counterfactual experiment should be, mainly, regarded as a device for out of sample evaluation of the model. An additional issue in the use of this procedure in policy evaluation, as noted in the literature, is that such a methodology is prone to the critique in Lucas (1976), as we are not considering the market reaction to the network change. Regarding those issues, we recall that endogenizing the network sequence is beyond the scope of the present investigation and that in Tonzer (2015), in a framework comparable to ours, exogeneity failed to be rejected by several tests. In our case, in addition, the time-varying nature of the network during the estimation period mitigates this issue. In fact, the market reaction to changes in the network should have been at least partially encoded in estimated parameters. Following this line of reasoning, we propose a new procedure to test the constancy of parameter estimates when we optimally change the network.

We estimate again the model using the in-sample Q periods and a $Q+1$ period in which we use $W^\star$ and the $\tilde{u}^\star$ of equation (22). We call the estimates for the network parameters obtained in this way $a^\star_{1,L}, a^\star_{1,R}, b^\star_{1,L}, b^\star_{1,R}$. We also estimate the model using the in-sample Q periods and the $Q+1$ realized out-of-sample period. Then, considering as given the estimates implied by the optimal network, we test for the joint hypothesis

$$H_0 : a_{1,L} = a^\star_{1,L}, a_{1,R} = a^\star_{1,R}, b_{1,L} = b^\star_{1,L}, b_{1,R} = b^\star_{1,R}, \tag{24}$$

using the estimates and covariance obtained by the realized sample in a Wald test. Failing to reject the null hypothesis would support parameter constancy and limit relevance of the Lucas critique, making possible the policy exercise Engle and Hendry (1993). In particular, we explicitly designed the Frobenius norm constraint equation (20) to increase the possibilities of obtaining parameter constancy. With those limitations in mind, we highlight that the product of our counterfactual exercise is a first step toward the determination of the target exposures that can be helpful for policy makers from a monitoring perspective.

3 Risk Spillovers among European Sovereign CDS

To better clarify the advantages and potential benefits of our methodology, we consider an application to the Euro area sovereign CDS premia. We restrict our attention to the Euro area because there is a documented currency firewall effect both for sovereign, Groba et al. (2013), and banks CDS, Alemany et al. (2015). Inside the EMU, we pick the two major economies and the peripheral countries excluding Greece, whose bailout make the CDS series unmanageable from the second semester of 2010. We use two different data sources: (i) the changes in the five-years sovereign CDS spreads for a selection of European countries and (ii) the matrices of foreign claims collected by the BIS. These data refer to the claims that the banking sector of a country A has with respect to the banking (public and private) sector of another country B. There is clearly an asymmetry between the dependent variable and the data source for the weighting matrices. Nevertheless, several theoretical models show tight linkages and feedback loops between sovereign and banking risk both in single and multi-country economies; see Bolton and Jeanne (2011); Acharya et al. (2014); Gennaioli et al. (2014); Farhi and Tirole (2016). In Dungey et al. (2017) this theoretical link is also empirically investigated. The aim of our analysis is to characterize, identify and evaluate the sovereign risk of the system, considering the total sovereign risk of the Euro area as the volatility of a weighted-average portfolio of European sovereign bonds. Risk spillovers are driven by the weight matrices, based on cross-country cross-credit exposures.

3.1 Data Description

3.1.1 Sovereign CDS spreads

We use the daily changes in the five-year sovereign CDS spreads, from 9/10/2008 to 30/12/2016, for France, Germany, Ireland, Italy, Portugal, and Spain, as downloaded from Thomson Reuters Eikon. As can be seen from Table U.1 in the supplementary material, mean and median are negligible, justifying our volatility approach. The Kurtosis of some series — in particular, that of Portugal, but also those of Germany and Ireland — is striking, adding a further motivation for a GARCH-type modeling.

The correlation is high between specific pairs — namely, Spain and Portugal, Spain and Italy, and Italy and Portugal — highlighting the closeness between those economies. Despite all being positive, several correlations display relatively small values, even if they are all significantly different from zero with a p-value below 0.01. Most interestingly, the smallest correlations are those between Germany and the other European countries (France excluded).

3.1.2 BIS Banking Statistics

We use data at a quarterly frequency to describe the network of foreign claims among Portugal, Italy, Ireland, Spain, France, and Germany from Q4 2008 to Q4 2016, as they are produced by BIS in the consolidated banking statistics (ultimate risk basis) [7]. BIS-consolidated banking statistics provide internationally comparable measures of national banking systems' exposures to country risk (see McGuire and Wooldridge (2005)). Only assets are reported. The residence of the ultimate obligor, or the country of ultimate risk, is defined as the country in which the guarantor of a financial claim resides, or the place in which the head office of a legally dependent branch is located. Our choice differs from that in Tonzer (2015) who, mainly motivated by sample-length consideration and exchange-rates-adjustment reasons, uses BIS locational banking statistics, in which the

[7] The quarterly claims are converted to a daily basis by repeating them for each day in the quarter. This choice assumes that claims variation is slower than changes in CDS variation. We also used a linear interpolation scheme, and the estimation results were unaltered.

residence of the obligor is the one of the local branch, even if owned abroad. We motivate this different choice by the importance of the local banking system in international financial intermediation (McCauley et al. (2010)), that could obfuscate the real flow of risk. Further, the obstacles, in Tonzer (2015) are not relevant in our framework because we focus on a smaller and more homogeneous group of countries, with a common currency, for which the more reliable consolidated-ultimate-risk-basis statistics are available. We expect that, if A reports a claim with B as a counterpart (i.e., if A's banking system owns a certain amount of B's public and private debt, investors will perceive A's sovereign risk to be dependent on B's sovereign risk in terms proportional to the claim amount.

We discuss now the impact of normalization schemes on the estimations. Taking into account the time variation for the spatial proximity matrices W_t obliges us to pay particular attention to the way in which we normalize these matrices. In fact, a simple row normalization at each time would make the comparison of the proximity matrices over time very difficult. Furthermore, a time-specific or matrix-specific normalization would lead to a loss of information, as both disregard the evolution over time of the network structure. The flexibility of our normalization schemes, is a consequence of of our careful treatment of identification and estimator properties.

In order to obtain parameters of a reasonable magnitude, but also to retain differences in matrix norms across time (which could be an important driver of dependence), we divide each row of W_t by an (economic) measure of the magnitude of the entities, which we denote as $M_{i,t}$. Our choice for the normalization M_{jt} of the j-th reporting country is its quarterly time series of total ultimate-risk-basis claims, which includes claims from the selected countries, but also from the rest of the world.[8] In addition, we stress that using this normalization allows us also to control for claims outside the chosen countries and to evaluate the exposure channel importance at the country level.

[8]We also investigated other choices for normalization: the absence of normalization, row normalization, the GDP of the reporting country, and the public debt of the reporting country. In the full sample estimation, total claims outperform, in likelihood terms, the alternative normalization schemes in the vast majority of models, and when this is not the case, the difference in likelihoods is negligible. Estimation results for these alternatives are available upon request.

[Insert figure 3 around here]

[Insert figure 4 around here]

The comparison between un-normalized series (Figure 3) and normalized series (Figure 4) show relevant differences. With respect to the un-normalized series, the deleveraging trend is less evident and the peripheral countries are more important. Thanks to the flexibility of our model we are able to use normalization as a device to include rest of the world effects and remove a deleveraging trend.

3.2 Parameter Estimation

In Table 3, we report QMLE results for the relevant parameters of the model. Table 2 includes some specification tests. We estimate the model by using a numerical-constrained optimization with bounds for the parameters function of the JSR of the network sequence. We checked covariance stationarity after the estimation by verifying the JSR condition of Theorem 1[9].

[Insert table 3 around here]

[Insert table 2 around here]

As the first panel of Table 2 shows, the SB-BEKK models outperform the diagonal BEKK model that does not include the networks, and it is the most used parsimonious restriction of a BEKK model. We check the joint significance of parameters both by likelihood ratio test statistics and a Wald-type statistic. Notably, both tests strongly reject the null, thus supporting the relevance of networks in variance-spillover analysis.

In addition, according to a Wald-type tests both right and left multiplication parameters have to be included in the analysis. This means that the Caporin and Paruolo (2015) model would have been misspecified in our empirical application. This is not surprising

[9] We compute a numerical bound obtained with the conic ellipsoid algorithm of the JSR matlab toolbox (see Vankeerberghen et al. (2014)).

giving that financial claims represent an asymmetric relationship and the use of left parameters would have focused only on risk receivers, neglecting the sources, underlined by right parameters and vice versa. Analogously remaining tests show that no part of the model can be excluded. Indeed, table 3 shows the significance of the vast majority of coefficients of the proposed model. In addition, the presence of significant coefficients of both signs induces variance-reducing contributions mediated by the claim matrix.

3.3 Inferred Networks

We use the test described at the beginning of subsection 2.1.3 to build a graphical representation that allows us to monitor the edges of the network — that is, the level of spillover between two specific countries as measured by the time-varying off-diagonal element of matrices A and B. In particular, one of the main features of our model is the ability to test the significance of the edges with a linear restriction (c.f. equation (6)). Due to the high number of coefficients and their time variation, it is difficult to conceive a tabular representation for our results. Instead, in order to display the topology of spillover flows, we propose a signed weighted directed network representation of the sequences of off diagonal elements in $A(W_t)$ and $B(W_t)$, with the weight proportional to the edge width and color representing the sign.

[Insert figure 5 around here]

In particular, in Figure 5, we report the relevant networks for three crisis periods: Q4 2008 (i.e., the collapse of Lehman Brothers and the bank bailout in Ireland), Q2 2010 (i.e., Greece's bailout), Q2 2016 (i.e., the Brexit referendum).(Section U in the supplementary material includes the complete representation.) Since we are conducting $2*(n^2-n)=60$ test, it could be argued that we need to adopt a multiple-hypothesis testing framework; for this reason, we apply a Bonferroni correction to the nominal 5% significance level and show in Figure 5 only edges with p-values smaller than $\frac{0.05}{60}=8.3e-04.$[10]. The similarity

[10]We report in Table U.2 of the supplementary material the p-values relevant for each off-diagonal

of pictures of different crisis periods is a consequence of our model that imposes a stable pattern on the network topology and signature. What could change is the strength of the linkage due to the exogenous change of the exposures.

The ARCH spillover topology shows how Spain, in the short term, is the central node receiving risk from Italy and Portugal and giving it to Ireland and France. Further, while France is giving risk back to Spain, Ireland is transferring risk to France and fueling a feedback loop with Germany. The overall picture is more involved in the GARCH part, with several feedback loops, the strongest of which is between Portugal and Spain. We register, also, the appearance of a feedback loop between the two major economies: Germany and France. Here, Ireland has the same incoming interaction, but has two new targets, Italy and Portugal. Regarding the signature, a negative sign implies a sign change of the incoming shock, but we remind the reader that covariance contribution is a quadratic form in the shocks, and therefore negative contribution does not directly imply a covariance reduction. Again, the presence of edges of differing signs in both series of matrices favors the presence of covariance reducing terms. However, given the non-linearity of the model, it is again difficult to understand where covariance reduction is located. In the next subsection, we will have a clearer description using covariance decompositions. The previous results point out the role of Spain and, to a lesser extent, of Ireland as risk middlemen able to transfer risk from peripheral countries to major economies.

3.4 Covariance Decompositions

The use of a Bonferroni correction could hide the role of less-significant but relevant links. In addition, it is difficult to understand where and when diversification benefits are at work. For these reasons, we use herein the methodology of Subsection 2.2 on the covariance decomposition of the SB-BEKK model. This tool will allow us to understand who benefited the most from variance-reducing contributions mediated by the network,

element of the sequences $A(W_t)$ and $B(W_t)$, which under our hypothesis of no measurement errors in weights is time invariant.

who is most impacted by network exposures, and who are the major risk spreaders.

[Insert figure 6 around here]

Figure 6 reports the percentage of the system-variance, constant, mixed, and indirect contributions.[11] The representation of the system is an equally weighted portfolio.[12] The mixed contribution has both variance-increasing and variance-reducing contributions, with only a few large peaks of diversifying effect, mostly around turmoil. The indirect contribution is always positive and present during turbulent periods. In the second part of the sample, the relevance of network contribution is reduced in favor of the constant, with the notable exception of some important episodes, such as the Brexit referendum. This change in behavior could be understood by the developments in the second half of 2012, with Mario Draghi's "whatever it takes" speech in July, which led to the Outright Monetary Transactions framework — but also with the July establishment and September implementation of the European Stability Mechanism. These surrogates of a centralized lender of last resort dampened the relevance of cross-border claims as a contagion channel. We summarize our findings with the following table 4, in which we report the cumulative percentage contribution on the whole estimation period.

[Insert table 4 around here]

Those results testify that most of the variance contribution is not mediated by the banking system exposures. This is not surprising since this is only one of several contagion channels. We also note how the indirect contribution is always superseded, in absolute value, by the mixed one, but the order of magnitude is, in the majority of cases, the same. In addition, the table shows that only Spain, Germany, and, to a lesser extent,

[11] We do not include the direct contribution since, although it is the biggest, it does not depend on the network links.

[12] It could be argued that an equally weighted portfolio does not represent a well-diversified portfolio. Indeed, this is the point of view put forward in Brunnermeier et al. (2017), which proposed weights proportional to the GDP to properly capture the sovereign credit risk of the Euro area. This possibility and the relevance of our model in evaluating the proposed European Safe Bonds will be investigated elsewhere.

Italy benefited from banking network effects. For Portugal and France in particular, and Ireland to a lesser extent, participation in the network of claims was dangerous. Looking again at the previously derived topology of spillover, we could now argue that those are the countries receiving risk from Spain and, in particular, that the countries most severely hit are the ones involved in an unbalanced feedback loop that favors Spain.

[Insert table 5 around here]

In Table 5. we report the downside. In fact, cumulative marginal spillover contributions, in percentage of the system variance, are able to identify the major sources of spillovers. Here we see that the most important countries are Portugal and Spain, but they contribute mostly in the amplification of their own variance. Instead, if we consider the contribution mediated through the banking exposures, the major sources of risk are Italy and Ireland, and, to a lesser extent, Spain. From the topology displayed before, we could argue that Ireland has a worrisome relationship with Germany, and that it is contributing to the risk of France, Portugal, and Italy. It seems that Italy's role as one of the major risk contributor to Spain has a big secondary impact to other countries through the previously described feedback loops.

3.5 Estimated Counterfactual Exposures

In the previous section, the bank exposure network is characterized as significant but minor in the sample spillover channel. In this subsection, we evaluate to what extent changing exposures could be beneficial by a counterfactual analysis, . As remarked in section 2.3 such alternative out of sample evaluation scheme is also due to the difficulty of retrieving intra-day sovereign CDS data, for building reliable volatility proxies.

We minimize the predicted path of the conditional average covariance in the system, looking for the optimal network structure, by following the methodology outlined in Subsection 2.3. We choose to optimize the forecast path in the quarter of the Brexit referendum (Q2 2016), while also including part of the subprime crisis and Greek bailout in the estimation sample (from Q4 2008 to Q1 2016). Our results convince us that the spillovers are channelled through foreign claims in the same way on the three occasions. We start by analyzing the counterfactual effect on the variance proxy of the equally weighted index. We perform the analysis by replacing the network of Q2 2016 with one of the following choices: a network with all exposures equal to zero, a network resulting from unconstrained optimization, a network coming from a redistribution-constrained optimization, and, finally, a network coming from a Frobenius-norm-constrained optimization. For the redistribution-constrained model, equation (18) implies that there is only a redistribution of the claims among the considered countries; for the Frobenius-constrained model, equation (20) implies that the change in the network is comparable with the historically registered changes.

[Insert figure7 around here]

Figure 7 shows the realized and counterfactual variance proxy of the equally weighted index during the sovereign debt crisis. We compute the variance according to equations (22) and (23). In addition, we plot the proxy coming from a network with zero exposures. The prevalent diversifying effect of exposures is evident in comparing the realized proxy with the zeroed one. The percentage change of the cumulative zeroed proxy with respect to the realized one is an increment of 16.55%. This also clarifies the fact that the naïve complete deleveraging solution, in our example, amplifies risk. This points out the necessity of a nontrivial exposure change, if we want to achieve risk mitigation. Our optimal counterfactual exercise is the first step in this direction. In particular, we think that several additional insights could be obtained by looking at the optimal proxy in 7. First of all, we are indeed able to reduce the out-of-sample counterfactual variance of the system by optimizing the network, based only on ex-ante information. This remains true also if we are imposing that the amount of claims for each country remain the same, or that the change is aligned with historical ones. As we were expecting, the optimization without constraints yielded the best percentage of cumulative reduction (-22.35%) and, in this case, the Brexit peak is mitigated and delayed. The drawback of this approach is that the variance is at several points in time above the realized one. On the opposite side, the reduction is halved, with the Frobenius-norm constraint (-11.18%), reduced by a lesser but comparable magnitude (-9.66%) with the redistribution constraint. The Brexit peak drop is smaller, but is indeed present and again higher for the Frobenius-norm constraint. Nevertheless, risk mitigation is much more stable in time and the variance is always less than or equal to the realized one in both constrained cases.[13]

[Insert table 6 around here]

The remarkable performance of the unconstrained procedure is obtained through drastic changes in the exposures (see Table U.3 in the supplementary material). We cannot be sure that extraordinary changes do not trigger a reaction from economic agents, which in turn could lead to a structural break in network parameters. As can be seen from our test on parameter constancy (see Table 6), we cannot reject the null hypothesis (i.e., parameter constancy) in this extreme case. In any case, we have stronger evidence considering the constrained instances. In particular, it is reassuring that the Frobenius-norm constraint has the lowest P-value, because it was designed as a solution to this particular issue. It is also interesting that an economically motivated constraint such as the redistribution one has comparable evidence in terms of parameter constancy. Considering the amount of variance reduction, its path through time, and the evidence for parameter constancy, the best compromise appears to be the optimal networks obtained by imposing the Frobenius-norm constraint.

[Insert table 7 around here]

To evaluate the feasibility of its changes, in Table 7, we report, for the Frobenius case only, the differences in billions of US dollars in the amounts needed to achieve the optimal network. Looking at Table 7, we rediscover some previous results. Here, the target

[13] Additionally, our procedure reduces an average correlation proxy (see Figure U.1 in the supplementary material). Instead, a procedure relying directly upon the minimization of conditional average correlation has the drawback of increasing the variance of some countries. Result are available upon request.

network proposes a better balance in the feedback loop with Portugal, and a deleverage in all the others. Furthermore, it is suggested that Germany, and to a lesser extent France, support the country, with other peripheral countries that divert their exposures from it. In general, the variance reduction we obtain is implied by exposures changes that would be hard to enforce in a single quarter. In our opinion, a lower but still meaningful variance reduction can be obtained by considering stricter and economically sound maximum redistribution constraints, leading to an implementable enforcement of redistribution. This is already possible with a minor modification of our methodology that enables us to account for any kind of constraint by simply changing the equation (18). In addition, the procedure is based on conditional covariance and could be regularly updated, closely following actual market evolution after the changes.

4 Conclusions

This paper illustrates how financial networks can be efficiently integrated within a multivariate GARCH model for risk analyses. We refer to the proposed framework as spatiotemporal econometrics of network-mediated risk, since it exploits spatial econometrics methodology in the investigation of a network describing risk relationships. Our spatiotemporal econometrics of risk enables a number of evaluations and analyses aimed at disentangling and understanding the role of asset interconnection in the evolution of the risk of the system. In particular, the new parsimonious econometric model we propose, the Spatial Bilateral BEKK, directly measures how variance and covariance spillovers flow through an exposure channel. Moreover, it entails the possibility of diversifying or enhancing spillover contributions through coefficients of both signs. Our approach builds on the introduction of spatial methods into volatility models as in Caporin and Paruolo (2015) and enhances the ability of proximity matrices to convey economic distances among assets and capture the interdependence across variables. This simple network model of an economic risk medium, based on data, allows a better investigation of the determinants of dependence and spillover effects, with respect to a purely statistical device.

From a methodological point of view, we make a number of contributions that go beyond the framework of Caporin and Paruolo (2015). We take advantage of the non-commutativity of adjacency matrices, and focus on both the risk-receiving propensity and risk-spreading effectiveness of spillovers. We use algebraic connectivity results in the study of identification and generalize previous results. We make use of the joint spectral radius for a set of matrices and of a VARMA representation of the model with time-varying coefficients to derive the convergence of a heterogeneous geometric series and to obtain covariance stationarity. Moreover, we derive the asymptotic normality of the QMLE under conditions that are standard in the literature. Again, the usefulness of our JSR approach is evident in adapting proof of consistency and asymptotic normality from the standard framework in Hafner and Preminger (2009). Finally, the study of the identification, covariance stationarity, and asymptotic normality in the general case allows us to include a general normalization procedure for proximity matrices, driven by economic insights, whose payoff is made clear in the empirical application.

The empirical application shows the ability of our model to give a reasonable description of European spillovers during the sovereign crisis, uncovering the fundamental role of France and Portugal as risk receivers and the risk-spreading effectiveness of Italy and Ireland. Our inferred network methodology also points out Spain's middleman role, which is not understandable from the single-parameter significance. We also derive a covariance decomposition that allows us to understand the network-mediated contribution to variance, pointing out in particular how several network configurations can reduce the covariance. This effect is possible only with our complete study of identification conditions, which leave unrestricted the sign of most of the model coefficients. In our empirical analysis, this diversifying effect was produced only by the mixed contribution in specific time periods. In addition, we document a change of relevance of exposures after regulatory interventions in the second half of 2012. Spillover spreading and receiving effectiveness are heterogeneous across countries, and we study them through our variance decomposition and the marginal spillover contribution. The picture that globally emerges, from our in-sample analysis, points to Ireland and Italy as the major sources of risk and Portugal and France as the recipients of most of it. However, while Portugal is also a lesser source of risk, shocks coming from France appear mostly innocuous, if not beneficial, having a small risk-reducing and stabilizing effect. Germany transforms most risks it receives into diversification benefits, while contributing negligibly as a risk spreader. Spain's less-trivial role exploits network exposures for diversification benefits, while at the same time it is the third source of network-mediated risk in the system. This middleman behavior in transferring risk from peripheral to major economies is direct with respect to France, and aided and mediated through Ireland with respect to Germany.

Finally, inspired by the monetary policy literature, we propose an alternative, policy oriented scheme for out of sample evaluation of the model based on counterfactual simulations. Our counterfactual analysis allows us to obtain target exposures for risk mitigation based only on ex-ante information. The empirical results on the quarter during which the Brexit referendum took place are an additional soundness check of our model. Results confirm the narrative implied by the in-sample analysis and the essential diversifying effect of banking-system exposures. Since the seminal works of Allen and Gale (2000), Eisenberg and Noe (2001) Freixas et al. (2000), networks of bank exposures were considered an important theoretical channel of spillover effects. The focus on covariance allows, in our opinion, a better description of the phenomenon with respect to the use of plain spatial models, already fruitful in this respect (see Tonzer (2015)). According to the recent review by Toniolo and White (2016) that focuses on the financial-stability mandate across countries and across history, the principal interventions that central banks took to maintain financial stability were the liquidity provision and the monitoring of systemically important financial institutions. We propose a new econometric framework with the ability to help the regulator to fulfil the financial system's monitoring requirement in an empirically measurable way. To achieve this ambitious goal, two major obstacles remain. The first is data disclosure: the bilateral exposures are made available only at the country level. The second one is methodological. In fact, the application of our methodology to a greater number of players requires the study of a covariance-targeting estimator of the model in order to obtain a number of parameters growing linearly with the number of nodes. Since rigorous covariance-targeting results for the standard multivariate volatility models were obtained only recently in the literature (Pedersen and Rahbek (2014), Francq et al. (2016)), this aspect is left for future research.

References

Viral Acharya, Itamar Drechsler, and Philipp Schnabl. A pyrrhic victory? bank bailouts and sovereign credit risk. *The Journal of Finance*, 69(6):2689–2739, 2014.

Aida Alemany, Laura Ballester, and Ana Gonzalez-Urteaga. Volatility spillovers in the european bank cds market. *Finance Research Letters*, 13:137 – 147, 2015.

Franklin Allen and Douglas Gale. Financial contagion. *Journal of Political Economy*, 108(1): pp. 1–33, 2000.

L. Anselin. *Spatial Econometrics: Methods and Models*. Studies in Operational Regional Science. Springer Netherlands, 2013.

Marco Avarucci, Eric Beutner, and Paolo Zaffaroni. On moment conditions for quasi-maximum likelihood estimation of multivariate arch models. *Econometric Theory*, 29(3):545–566, 2013.

Geert Bekaert and Campbell R Harvey. Emerging equity market volatility. *Journal of Financial economics*, 43(1):29–77, 1997.

Ruslan Bikbov and Mikhail Chernov. Monetary policy regimes and the term structure of interest rates. *Journal of Econometrics*, 174(1):27 – 43, 2013.

Monica Billio and Loriana Pelizzon. Volatility and shocks spillover before and after emu in european stock markets. *Journal of Multinational Financial Management*, 13(4):323–340, 2003.

Monica Billio, Massimiliano Caporin, Roberto Calogero Panzica, and Loriana Pelizzon. The impact of network connectivity on factor exposures, asset pricing and portfolio diversification. SAFE Working Paper Series 166, Research Center SAFE - Sustainable Architecture for Finance in Europe, Goethe University Frankfurt, 2017.

Francisco Blasques, Siem Jan Koopman, Andre Lucas, and Julia Schaumburg. Spillover dynamics for systemic risk measurement using spatial financial time series models. *Journal of Econometrics*, 195(2):211 – 223, 2016.

Vincent D. Blondel and Yurii Nesterov. Computationally efficient approximations of the joint spectral radius. *SIAM Journal on Matrix Analysis and Applications*, 27(1):256–272, 2005.

Patrick Bolton and Olivier Jeanne. Sovereign default risk and bank fragility in financially integrated economies. *IMF Economic Review*, 59(2):162–194, Jun 2011.

Philippe Bougerol and Nico Picard. Strict stationarity of generalized autoregressive processes. *Ann. Probab.*, 20(4):1714–1730, 10 1992.

Farid Boussama, Florian Fuchs, and Robert Stelzer. Stationarity and geometric ergodicity of bekk multivariate garch models. *Stochastic Processes and their Applications*, 121(10):2331–2360, 2011.

Andreas Brandt. The stochastic equation yn 1=anyn bn with stationary coefficients. *Advances in Applied Probability*, 18(1):211,Äì220, 1986.

Richard A. Brualdi. Kronecker products of fully indecomposable matrices and of ultrastrong digraphs. *Journal of Combinatorial Theory*, 2(2):135 – 139, 1967.

Markus K. Brunnermeier, Sam Langfield, Marco Pagano, Ricardo Reis, Stijn Van Nieuwerburgh, and Dimitri Vayanos. Esbies: safety in the tranches. *Economic Policy*, 32(90):175–219, 2017.

Massimiliano Caporin and Paolo Paruolo. Proximity-structured multivariate volatility models. *Econometric Reviews*, 34(5):559–593, 2015.

Giancarlo Corsetti, Marcello Pericoli, and Massimo Sbracia. Correlation analysis of financial contagion. *Financial contagion: The viral threat to the wealth of nations*, pages 11–20, 2011.

Mardi Dungey, John Harvey, and Vladimir Volkov. The changing international network of sovereign debt and financial institutions. Working Papers 2017-04, University of Tasmania, Tasmanian School of Business and Economics, 2017.

Larry Eisenberg and Thomas H Noe. Systemic risk in financial systems. *Management Science*, 47(2):236–249, 2001.

J. Paul Elhorst. Dynamic models in space and time. *Geographical Analysis*, 33(2):119–140, 2001.

J Paul Elhorst. Specification and estimation of spatial panel data models. *International regional science review*, 26(3):244–268, 2003.

Robert F. Engle and David F. Hendry. Testing superexogeneity and invariance in regression models. *Journal of Econometrics*, 56(1):119 – 139, 1993.

Robert F Engle and Kenneth F Kroner. Multivariate simultaneous generalized arch. *Econometric theory*, 11(01):122–150, 1995.

Emmanuel Farhi and Jean Tirole. Deadly embrace: Sovereign and financial balance sheets doom loops. Working Paper 21843, National Bureau of Economic Research, January 2016.

Trevor I. Fenner and Georghios Loizou. On fully indecomposable matrices. *Journal of Computer and System Sciences*, 5(6):607 – 622, 1971.

Miroslav Fiedler. Algebraic connectivity of graphs. *Czechoslovak mathematical journal*, 23(2): 298–305, 1973.

Miroslav Fiedler. A property of eigenvectors of nonnegative symmetric matrices and its application to graph theory. *Czechoslovak Mathematical Journal*, 25(4):619–633, 1975.

Kristin Forbes. The" big c": Identifying contagion. Technical report, National Bureau of Economic Research, 2012.

Christian Francq and Jean-Michel Zakoian. Comments on the paper by minxian yang: ,Äúsome properties of vector autoregressive processes with markov-switching coefficients,Äù. *Econometric Theory*, 18(3):815,Äì818, 2002. doi: 10.1017/S0266466602183125.

Christian Francq, Lajos Horvath, and Jean-Michel Zakoian. Variance targeting estimation of multivariate garch models. *Journal of Financial Econometrics*, 14(2):353–382, 2016.

Xavier Freixas, Bruno M Parigi, and Jean-Charles Rochet. Systemic risk, interbank relations, and liquidity provision by the central bank. *Journal of money, credit and banking*, pages 611–638, 2000.

Nicola Gennaioli, Alberto Martin, and Stefano Rossi. Sovereign default, domestic banks, and financial institutions. *The Journal of Finance*, 69(2):819–866, 2014.

Jonatan Groba, Juan A. Lafuente, and Pedro Serrano. The impact of distressed economies on the eu sovereign market. *Journal of Banking & Finance*, 37(7):2520 – 2532, 2013.

Leonid Gurvits. Stability of discrete linear inclusion. *Linear algebra and its applications*, 231: 47–85, 1995.

Christian M Hafner and Arie Preminger. On asymptotic theory for multivariate garch models. *Journal of Multivariate Analysis*, 100(9):2044–2054, 2009.

Christian M. Hafner and Jeroen V. K. Rombouts. Estimation of temporally aggregated multivariate garch models. *Journal of Statistical Computation and Simulation*, 77(8):629–650, 2007.

Raphaël Jungers. *The Joint Spectral Radius: Theory and Applications*. Lecture Notes in Control and Information Sciences. Springer Berlin Heidelberg, 2009. ISBN 9783540959793.

Sebastian Keiler and Armin Eder. Cds spreads and systemic risk: A spatial econometric approach. *Discussion Paper N.01/2013 Deutsche Bundesbank*, 2013.

Sebastien Laurent, Jeroen V. K. Rombouts, and Francesco Violante. On the forecasting accuracy of multivariate garch models. *Journal of Applied Econometrics*, 27(6):934–955, 2012.

Lung-fei Lee and Jihai Yu. Qml estimation of spatial dynamic panel data models with time varying spatial weights matrices. *Spatial Economic Analysis*, 7(1):31–74, 2012.

James LeSage and Robert Kelley Pace. *Introduction to spatial econometrics*. CRC press, 2009.

James P LeSage and R Kelley Pace. Interpreting spatial econometric models. In *Handbook of Regional Science*, pages 1535–1552. Springer, 2014.

Robert E. Lucas. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1(Supplement C):19 – 46, 1976.

J.R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley Series in Probability and Statistics: Texts and References Section. Wiley, 1999. ISBN 9780471986331.

Bennet T. Mccallum. Robustness properties of a rule for monetary policy. *Carnegie-Rochester Conference Series on Public Policy*, 29(Supplement C):173 – 203, 1988.

Robert McCauley, Patrick McGuire, and Goetz von Peter. The architecture of global banking: from international to multinational? *BIS Quarterly Review*, page 25, 2010.

Patrick McGuire and Philip Wooldridge. The bis consolidated banking statistics: structure, uses and recent enhancements. *BIS Quarterly Review*, page 73, 2005.

Angela Ng. Volatility spillover effects from japan and the us to the pacific–basin. *Journal of international money and finance*, 19(2):207–233, 2000.

Rasmus S. Pedersen and Anders Rahbek. Multivariate variance targeting in the bekk-garch model. *The Econometrics Journal*, 17(1):24–55, 2014.

Dimitris N Politis and Joseph P Romano. A circular block-resampling procedure for stationary data. *Exploring the limits of bootstrap*, pages 263–270, 1992.

Giorgio E. Primiceri. Time varying structural vector autoregressions and monetary policy. *The Review of Economic Studies*, 72(3):821–852, 2005.

Xi Qu, Lung fei Lee, and Jihai Yu. Qml estimation of spatial dynamic panel data models with endogenous time varying spatial weights matrices. *Journal of Econometrics*, 197(2):173 – 201, 2017.

Oscar Rojo. A nontrivial upper bound on the largest laplacian eigenvalue of weighted graphs. *Linear Algebra and its Applications*, 420(2):625 – 633, 2007.

Gian-Carlo Rota and W. Gilbert Strang. A note on the joint spectral radius. *Indagationes Mathematicae (Proceedings)*, 63:379 – 381, 1960.

Julio J. Rotemberg and Michael Woodford. An optimization-based econometric framework for the evaluation of monetary policy. *NBER Macroeconomics Annual*, 12:297–346, 1997.

R. Tarjan. Depth-first search and linear graph algorithms. *SIAM Journal on Computing*, 1(2): 146–160, 1972.

Gianni Toniolo and Eugene N White. The evolution of the financial stability mandate. In M.D. Bordo, Ø. Eitrheim, M. Flandreau, and J.F. Qvigstad, editors, *Central Banks at a Crossroads: What Can We Learn from History?*, Studies in Macroeconomic History, chapter 11, pages 424–492. Cambridge University Press, 2016.

Lena Tonzer. Cross-border interbank networks, banking risk and contagion. *Journal of Financial Stability*, 18:19 – 32, 2015.

Guillaume Vankeerberghen, Julien Hendrickx, and Raphaël M. Jungers. Jsr: A toolbox to compute the joint spectral radius. In *Proceedings of the 17th International Conference on*

Hybrid Systems: Computation and Control, HSCC '14, pages 151–156, New York, NY, USA, 2014. ACM. ISBN 978-1-4503-2732-9.

A Proofs and Auxiliary Results

A.1 Identification

A.1.1 Proof of Lemma 1 on page 11

Proof.

$$\begin{aligned}
M(W_t) &= \begin{bmatrix} I_n \otimes I_n & W_k' \otimes I_n & I_n \otimes W_k \end{bmatrix} (I_3 \otimes H) \\
&= \sum_{i=1}^{n} \left[(e_i \otimes e_i) e_i', (W_k' e_i \otimes e_i) e_i', (e_i \otimes W_k e_i) e_i' \right] \\
M(W_k)' M(W_k) &= \begin{bmatrix} I_n & \\ & \mathcal{M}(W_k) \end{bmatrix} \quad \mathcal{M}(W_k) = \begin{bmatrix} D_{W_k \odot W_k} & W_k \odot W_k \\ W_k' \odot W_k' & D_{W_k' \odot W_k'} \end{bmatrix} \\
D_{W_k' \odot W_k'} &= (W_k' \odot W_k') \mathbf{1}_n \quad , \quad D_{W_k \odot W_k} = (W_k \odot W_k) \mathbf{1}_n \ .
\end{aligned}$$

Having the identity I_n as one of the diagonal blocks, $M(W_k)' M(W_k)$ has n eigenvalues equal to one. In addition, using the property of the determinant for block matrices, the characteristic polynomial of $\mathcal{M}(W_k)$, is:

$$\det\left(\mathcal{M}(W_k) - \lambda I_{n^2}\right) = \det\left(D_{W_k^{\odot} W_k} - \lambda I_n\right) \det\left(L(\mathfrak{G}_k) - \lambda I_n\right) = 0$$

$\det\left(D_{W_k^{\odot} W_k} - \lambda I_n\right) = 0$ has exactly n solutions that can be read off as the degrees $d_{\mathfrak{G},i}$. The remaining eigenvalues are the eigenvalues of $L(\mathfrak{G}_k)$. Considering the second part of the theorem, it is well known since the work of Fiedler (Fiedler, 1973, 1975) that the number of null eigenvalues of $L(\mathfrak{G}_k)$ is the number of connected components c_k of $\mathfrak{G}_k$. If we have s_k single nodes without self-loops, then also s_k nodes have zero degree, resulting in additional s_k null eigenvalues of $M(W_k)' M(W_k)$. □

A.1.2 Proof of Proposition 1 on page 11

Proof. Without loss of generality, let us consider $A(W_k)$, we have:

$$\begin{aligned}
\operatorname{vec}(A(W_k)) &= \operatorname{vec}(\operatorname{diag}(a_0)) + W_k' \otimes I_n \operatorname{vec}(\operatorname{diag}(a_{1L})) + I_n \otimes W_k \operatorname{vec}(\operatorname{diag}(a_{1R})) \\
&= \begin{bmatrix} I_n \otimes I_n & W_k' \otimes I_n & I_n \otimes W_k \end{bmatrix} \begin{bmatrix} \operatorname{vec}(\operatorname{diag}(a_0)) \\ \operatorname{vec}(\operatorname{diag}(a_{1,L})) \\ \operatorname{vec}(\operatorname{diag}(a_{1,R})) \end{bmatrix} = M(W_k) \begin{bmatrix} a_0 \\ a_{1,L} \\ a_{1,R} \end{bmatrix} .
\end{aligned}$$

and analogously

$$\operatorname{vec}\left(\left[A(W_1)' : A(W_2)' : \ldots : A(W_K)'\right]'\right) = \left[M(W_1)' : M(W_2)' \ldots : M(W_K)'\right]' \begin{bmatrix} a_0 \\ a_{1,L} \\ a_{1,R} \end{bmatrix}$$

The $[Kn^2 \times 3n]$ matrix $\bar{M} = \left[(W_1)' : M(W_2)' \ldots : M(W_K)'\right]'$, is of column rank equal to $\max\limits_{1\ldots k} \operatorname{rank}(M(W_k))$. Given that the rank of a generic matrix T is equal to the rank of $T'T$, and lemma 1 on page 11 $M(W_k)$ is of column rank $3n - c_k - s_k$. Then, if we impose $\chi = \min\limits_{1\ldots k} c_k + s_k$ linear restriction on the parameter vector, variation in the parameters induces a unique variation in π_K. □

A.1.3 Proof of Lemma 2 on page 11

Proof. The result follow from the fact that the non zero elements of $\mathfrak{W}_k$ are the same of $(W_k' \odot W_k')(W_k \odot W_k)$ and consequently of $W_k' W_k$. Then, since the multiplication of two fully indecomposable matrices is fully indecomposable, and fully indecomposability implies irreducibility, Fenner and Loizou (1971), under the hypothesis of the lemma there is a at least one $k = 1 \ldots K$ for which $\mathfrak{W}_k$ has only one component. Then, the result follows from Proposition 1. □

A.1.4 Identification for Symmetric Matrices

Corollary 1. *If the matrices W_k are all symmetric, to achieve identification the model must include either the left multiplication or the right multiplication elements. The previous result holds with a corresponding simplification in the structure of $M(W_k)$ and the model is identified if the corresponding $\bar{M}$ is full column rank.*

Proof. The corollary is a consequence of symmetry of matrices W_k. Suppose we focus on the shock component and assume a constant W. We have $A_L u_{t-1} u_{t-1}' A_L' = (A_0 + A_{1,L} W) u_{t-1} u_{t-1}' (A_0 + W' A_{1,L})$ thanks to the diagonal form of the parameter matrices. Moreover, by symmetry, $(A_0 + A_{1,L} W) u_{t-1} u_{t-1}' (A_0 + W' A_{1,L})$ and the latter equals $(A_0 + W' A_{1,L}) u_{t-1} u_{t-1}' (A_0 + A_{1,L} W)$. The latter is equal to the right multiplication case if $W = W'$.

Then, with only left(right) parameters, we have :

$$M(W_k)' M(W_k) = \begin{bmatrix} I_n & \\ & D_{W_k \odot W_k} \end{bmatrix}.$$

This matrix is full rank unless some of the row(column) sums are zero. Since the W_k are non negative this can happen only if the row(column) elements are all zero and we reduce to the conditions given in (Caporin and Paruolo, 2015) □

A.2 Covariance Stationarity

The following lemma considers a deterministic network sequence and gives a sufficient condition on the joint spectral radius such that it probably collapses to the one given in terms of the top Lyapunov exponent in Brandt (1986) for scalar coefficients. Bougerol and Picard (1992) extended the last result to coefficient matrices.[14] Note that results in Bougerol and Picard (1992) are valid under the assumption that the network sequence has an i.i.d distrbution.

[14] A similar condition is also used by Hafner and Preminger (2009) in showing stationarity and ergodicity of the VEC model.

Lemma 3 (Convergence of Heterogeneous Geometric Series)**.** *If $\varrho(\mathcal{A}) < 1$ then*

$$S = \lim_{T\to\infty} S_n = \lim_{T\to\infty} \sum_{t=0}^{T} \prod_{i=0}^{t} A_i < \infty \tag{25}$$

Proof. We can use here the Cauchy convergence criterion. Given $m > n \in \mathcal{N}$ with fixed but arbitrary $r = m - n$, we have to show that $\|S_m - S_n\| \to 0$ when $n \to \infty$.

We have :

$$0 \le \|S_m - S_n\| = \left\|\sum_{k=n}^{m} B_k\right\| \le \sum_{k=n}^{m} \sup_{A_{\sigma_1},\dots,A_{\sigma_k}\in\mathcal{A}} \left(\left\|\prod_{i=1}^{k} A_{\sigma_i}\right\|\right) \tag{26}$$

and we can write

$$\begin{aligned}
\lim_{n\to\infty} \sum_{k=n}^{m} \sup_{\sup_{A_{\sigma_1},\dots,A_{\sigma_k}\in\mathcal{A}}} \left(\left\|\prod_{i=1}^{k} A_{\sigma_i}\right\|\right) &= \lim_{n\to\infty} \sum_{k=n}^{m} \left(\hat{\rho}_k(\mathcal{A})\right)^k \\
= \lim_{n\to\infty} \varrho(\mathcal{A})^n \lim_{n\to\infty} \sum_{k=n}^{m} \frac{\left(\hat{\rho}_k(\mathcal{A})\right)^k}{\varrho(\mathcal{A})^n} &= \lim_{n\to\infty} \frac{1-\varrho(\mathcal{A})^{r+1}}{1-\varrho(\mathcal{A})} \varrho(\mathcal{A})^n \to 0
\end{aligned}$$

□

The use of the joint spectral radius leads, also, to an interesting property for matrix sets that leave a proper cone invariant.

Definition A.1. *Cones and Proper Cones. A cone in $\mathbb{R}^n$ is a subset $K \subseteq \mathbb{R}^n$ such that $\lambda v \in K$ for all $\lambda \ge 0$ and $v \in K$. K is proper if it is closed, convex, has non empty interior, and contains no straight lines.*

Definition A.2. *Cone Invariance. $A_i \in \mathcal{A}$ leave a proper cone invariant if there exist a proper cone $K \subseteq \mathbb{R}^n$ such that if $v \in K$ $u_i = A_i v \in J_i$ with $J_i \subseteq K$ for each i.*

According to Blondel and Nesterov (2005) it is possible to define a norm $\|\cdot\|_K$ associated with the cone K such that, if A and B leave the proper cone K invariant, then $\|A\|_K \le \|A+B\|_K$. This norm will be the essential in the proof of the following lemma.

Lemma 4. *Consider two infinite set of $n \times n$ matrices $\mathcal{A} = \{A_i\}_{i=0}^{\infty}$ and $\mathcal{B} = \{B_i\}_{i=0}^{\infty}$ and their sum $\mathcal{A}+\mathcal{B} = \{M \in \mathcal{A}+\mathcal{B} | M = A_i + B_i, A_i \in A, B_i \in \mathcal{B}\}$ with $A_0, B_0 = I_n$. Suppose that $\mathcal{A}$ and $\mathcal{B}$ leave the cone K invariant. We have $\varrho(\mathcal{B}) \le \varrho(\mathcal{A}+\mathcal{B})$.*

Proof. $\prod_{i=1}^{t} A_{\sigma_i} + B_{\sigma_i}$ is a sum of 2^t products of t terms. By the assumptions the matrices C_j $j = 1, \dots, 2^t$ obtained by each of this products leave K invariant. In addition we note that $C_{2^t} = \prod_{i=1}^{t} B_{\sigma_i}$. We obtain

$$\left\|\prod_{i=1}^{t} (A_{\sigma_i} + B_{\sigma_i})\right\|_K = \left\|\prod_{i=1}^{t} B_{\sigma_i} + \sum_{j=1}^{2^t-1} C_j\right\|_K \ge \left\|\prod_{i=1}^{t} B_{\sigma_i}\right\|_K$$

so

$$\begin{aligned}\varrho(\mathcal{B}) &= \lim_{t\to\infty} \sup_{B_{\sigma_1},\dots,B_{\sigma_t}\in\mathcal{B}} \left(\left\|\prod_{i=1}^{t} B_{\sigma_i}\right\|_K\right)^{1/t} = \lim_{t\to\infty} \sup_{\left(A_{\sigma_1}+B_{\sigma_1}\right),\dots,\left(A_{\sigma_t}+B_{\sigma_t}\right)\in(\mathcal{A}+\mathcal{B})} \left(\left\|\prod_{i=1}^{t} B_{\sigma_i}\right\|_K\right)^{1/t} \\ &\leq \lim_{t\to\infty} \sup_{\left(A_{\sigma_1}+B_{\sigma_1}\right),\dots,\left(A_{\sigma_t}+B_{\sigma_t}\right)\in(\mathcal{A}+\mathcal{B})} \left(\left\|\prod_{i=1}^{t} \left(A_{\sigma_i}+B_{\sigma_i}\right)\right\|_K\right)^{1/t} = \varrho(\mathcal{A}+\mathcal{B})\end{aligned}$$

□

A.2.1 Proof of theorem 2 on page 16

Proof. Let us consider the filtration $\mathcal{F}_{t-1} = \{\mathcal{U}_{t-1}, \mathcal{W}_{t-1}\}$ where $\mathcal{U}_{t-1}$ is the information set given by the past u_t and $\mathcal{W}_{t-1}$ is the information set of the past network W_t. We remark that, by definition of the processes, $\mathcal{W}_{t-2} \subset \mathcal{U}_{t-1}$.

Since η_t is a martingale difference sequence $\mathbb{E}\left[\eta_t \left|\mathcal{U}_{t-1}, \mathcal{W}_{t-1}\right.\right] = 0$, and we also have

$$\mathbb{E}\left[X_t \left|\mathcal{U}_{t-1}, \mathcal{W}_{t-1}\right.\right] = \tilde{C} + \left(\tilde{A}(W_t) + \tilde{B}(W_t)\right) X_{t-1} - \tilde{B}(W_t)\,\eta_{t-1}.$$

Using the iterated expectation theorem and the recursion for X_t

$$\begin{aligned}\mathbb{E}\left[X_t \left|\mathcal{U}_{t-2}, \mathcal{W}_{t-1}\right.\right] &= \mathbb{E}\left[\mathbb{E}\left[X_t \left|\mathcal{U}_{t-1}, \mathcal{W}_{t-1}\right.\right] \left|\mathcal{U}_{t-1}, \mathcal{W}_{t-1}\right.\right] \\ &= \mathbb{E}\left[\tilde{C} + \left(\tilde{A}(W_t) + \tilde{B}(W_t)\right) X_{t-1} - \tilde{B}(W_t)\,\eta_{t-1} \left|\mathcal{U}_{t-2}, \mathcal{W}_{t-1}\right.\right] \\ &= \tilde{C} + \left(\tilde{A}(W_{t-1}) + \tilde{B}(W_{t-1})\right) \mathbb{E}\left[X_{t-1} \left|\mathcal{U}_{t-2}, \mathcal{W}_{t-1}\right.\right].\end{aligned}$$

Analogously,

$$\begin{aligned}\mathbb{E}\left[X_t \left|\mathcal{U}_{t-3}, \mathcal{W}_{t-1}\right.\right] &= \left[I + \left[\tilde{A}(W_{t-1}) + \tilde{B}(W_{t-1})\right]\right] \tilde{C} \\ &+ \left[\tilde{A}(W_{t-1}) + \tilde{B}(W_{t-1})\right] \left[\tilde{A}(W_{t-2}) + \tilde{B}(W_{t-2})\right] \mathbb{E}\left[X_{t-2} \left|\mathcal{U}_{t-3}, \mathcal{W}_{t-1}\right.\right],\end{aligned}$$

Further, by induction

$$\begin{aligned}\mathbb{E}\left[X_t \left|\mathcal{U}_{t-\tau}, \mathcal{W}_{t-1}\right.\right] &= \left\{I + \sum_{k=1}^{\tau-2} \prod_{l=1}^{k} \left[\tilde{A}(W_{t-l}) + \tilde{B}(W_{t-l})\right]\right\} \tilde{C} \\ &+ \left\{\prod_{l=1}^{\tau-1} \left[\tilde{A}(W_{t-l}) + \tilde{B}(W_{t-l})\right]\right\} \mathbb{E}\left[X_{t-\tau+1} \left|\mathcal{U}_{t-\tau}, \mathcal{W}_{t-1}\right.\right].\end{aligned}$$

When $\tau \to \infty$, if $\varrho\left(\left\{\tilde{A}(W_t) + \tilde{B}(W_t)\right\}_{t=-\infty}^{\infty}\right) < 1$, by Lemma 3, for all possible network sequences, the first term converges to a finite limit and the remainder goes to zero according to Gurvits (1995) and Jungers (2009). □

A.3 Consistency and Asymptotic Normality

In this section we denote as $\|\cdot\|$ the operator norm. Different norms are specified when needed. D_n is the duplication matrix and C_{nm} is the commutation matrix. Under assumption 1 and by Lemma 4, since the set of half vectorized symmetric positive semidefinite matrices is a proper cone of the set of half vectorized symmetric matrices, i.e of the vector space $\mathbb{R}^{(n(n+1)/2)}$, and $\tilde{A}(W_t)$ and $\tilde{B}(W_t)$ leave this cone invariant (c.f subsection 4.2 in Boussama et al. (2011)). we have

$$\varrho_B = \varrho\left(\left\{\tilde{B}(W_t)\right\}_{t=-\infty}^{\infty}\right) \leq \varrho\left(\left\{\tilde{A}(W_t) + \tilde{B}(W_t)\right\}_{t=-\infty}^{\infty}\right) < 1 \quad , \quad \bar{\varrho}_B = \sup_{\theta\in\Theta} \varrho_B < 1.$$

Lemma 5. *Under Assumptions 1-4* $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\Sigma_t - \tilde{\Sigma}_t\right\|^{s/2}\right] = O\left(\bar{\varrho}_B\right)$

Proof. By solving equation (5) recursively, we get

$$\begin{aligned}
\xi_t &= \mathrm{vech}\left(\Sigma_t\right) = \tilde{C} + \tilde{A}(W_t) X_{t-1} + \tilde{B}(W_t)\xi_{t-1} \\
&= \tilde{C} + \tilde{A}(W_t) y_{t-1} + \sum_{\tau=1}^{t-1}\left\{\prod_{l=1}^{\tau-1}\tilde{B}(W_{t-l})\right\}\left(\tilde{C} + \tilde{A}(W_{t-\tau}) X_{t-\tau-1}\right) + \left\{\prod_{l=1}^{t-1}\tilde{B}(W_{t-l})\right\}\xi_0.
\end{aligned}$$

In addition

$$\left\|\Sigma_t - \tilde{\Sigma}_t\right\| \leq \left\|\Sigma_t - \tilde{\Sigma}_t\right\|_2 \leq \left\|D_n^+\right\|\left\|D_n\right\|\left\|\xi_t - \tilde{\xi}_t\right\| \leq \left\|D_n^+\right\|\left\|D_n\right\|\varrho_B^t\left\|\xi_0 - \tilde{\xi}_0\right\|$$

By Assumption 2 and 3, and the c_r inequality we have $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\Sigma_t - \tilde{\Sigma}_t\right\|^{s/2}\right] = O\left(\bar{\varrho}_B\right)$

□

Lemma 6. *Under Assumptions 1-7*

$$i)\quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial\tilde{\xi}_t}{\partial a'}\right\|^3\right] < \infty \qquad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial\tilde{\xi}_t}{\partial b'}\right\|^3\right] < \infty$$

$$ii)\quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2\tilde{\xi}_t}{\partial a'\partial a'}\right\|^3\right] < \infty \quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2\tilde{\xi}_t}{\partial b'\partial a'}\right\|^3\right] < \infty \quad \mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\frac{\partial^2\tilde{\xi}_t}{\partial b'\partial b'}\right\|^3\right] < \infty$$

Proof. Using the linear map that links $A(W_t)$ and $B(W_t)$ to a and b, it is possible to obtain the derivatives of $\tilde{A}(W_t)$ and $\tilde{B}(W_t)$. Consider, as an example, $\tilde{B}(W_t)$:

$$
\begin{aligned}
\operatorname{vec}\left(\tilde{B}(W_t)\right) &= \left(D_n' \otimes D_n^+\right)(I_n \otimes C_{nn} \otimes I_n)\operatorname{vec}(M(W_t)b) \otimes \operatorname{vec}(M(W_t)b) \\
&= \left(D_n' \otimes D_n^+\right)(I_n \otimes C_{nn} \otimes I_n)(M(W_t) \otimes M(W_t))(b \otimes b) \\
\frac{\partial}{\partial b'}\operatorname{vec}\left(\tilde{B}(W_t)\right) &= (D_n' \otimes D_n^+)(I_n \otimes C_{nn} \otimes I_n)(M(W_t) \otimes M(W_t))\, 2D_{3n}D_{3n}^+(b \otimes I_{3n}) \\
\frac{\partial}{\partial b'}\operatorname{vec}\left(\frac{\partial}{\partial b'}\operatorname{vec}\left(\tilde{B}(W_t)\right)\right) &= \left[I_{3n} \otimes \left[(D_n' \otimes D_n^+)(I_n \otimes C_{nn} \otimes I_n)(M(W_t) \otimes M(W_t))\, 2D_{3n}D_{3n}^+\right]\right]\frac{\partial}{\partial b'}\operatorname{vec}(b \otimes I_{3n}) \\
&= \left[I_{3n} \otimes \left[(D_n' \otimes D_n^+)(I_n \otimes C_{nn} \otimes I_n)(M(W_t) \otimes M(W_t))\, 2D_{3n}D_{3n}^+\right]\right] \times \\
&\times \; (1 \otimes C_{3n3n} \otimes I_{3n})(I_{3n} \otimes \operatorname{vec}(I_{3n}))
\end{aligned}
$$

Let us define the following quantities

$$
\begin{array}{ccc}
\|C_{nn}\| = K_{1,n} & \|(D_n' \otimes D_n^+)\| = K_{2,n} & \|(D_n D_n^+)\| = K_{3,n} \\
\sup_{\theta\in\Theta}\|a\| = K_a & \sup_{\theta\in\Theta}\|b\| = K_b & \sup_{\theta\in\Theta}\|C\| = K_c.
\end{array}
$$

Then, let us bound

$$
\begin{aligned}
\|(M(W_t) \otimes M(W_t))\| &= \sqrt{\rho\left((M(W_t) \otimes M(W_t))'(M(W_t) \otimes M(W_t))\right)} \\
&= \rho\left(M(W_t)' M(W_t)\right) = \rho_M.
\end{aligned}
$$

We already studied the characteristic polynomial in lemma 1 on page 11 ; it has n eigenvalues equal to 1, eigenvalues equal to $d_{\mathfrak{G},i}$ for $i = 1, \ldots, n$ and the last n eigenvalues are the eigenvalues of the Laplacian $L(\mathfrak{G}_t)$. In particular, we are interested in $\rho(L(\mathfrak{G}_t))$. Bounds could be found in Rojo (2007), the most trivial one being the column sum over row maxima of the weight matrix.

$$
\rho(L(\mathfrak{G}_t)) < \sum_{j=1}^{n} \max_{i=1,\ldots,n} \mathfrak{W}_{t,ij}
$$

By Assumption 7 we can bound $W_{t,ij} \leq d^*$ and $d_{\mathfrak{G},i} \leq n\left(d^*\right)^2$. We have

$$\begin{aligned}
\rho_M = \max\left(1, n\left(d^*\right)^2, \sum_{j=1}^{n} \max_{i=1,\dots,n} \mathfrak{W}_{t,ij}\right) &\leq \max\left(1, n\left(d^*\right)^2\right) && (27)\\
\sup_{\theta\in\Theta, t\in(-\infty,\infty)} \left\|\mathrm{vec}\left(\tilde{B}\left(W_t\right)\right)\right\| &\leq K_{2,n}K_{1,n}\rho_M K_b^2 && (28)\\
\sup_{\theta\in\Theta, t\in(-\infty,\infty)} \left\|\frac{\partial}{\partial b'}\mathrm{vec}\left(\tilde{B}\left(W_t\right)\right)\right\| &\leq K_{2,n}K_{1,n}\rho_M 2K_{3,3n}K_b && (29)\\
\sup_{\theta\in\Theta, t\in(-\infty,\infty)} \left\|\frac{\partial}{\partial b'}\mathrm{vec}\left(\frac{\partial}{\partial b'}\mathrm{vec}\left(\tilde{B}\left(W_t\right)\right)\right)\right\| &\leq K_{2,n}K_{1,n}\rho_M 2K_{3,3n}K_{1,3n} && (30)\\
\sup_{\theta\in\Theta, t\in(-\infty,\infty)} \left\|\mathrm{vec}\left(\tilde{A}\left(W_t\right)\right)\right\| &\leq K_{2,n}K_{1,n}\rho_M K_b^2 && (31)\\
\sup_{\theta\in\Theta, t\in(-\infty,\infty)} \left\|\frac{\partial}{\partial a'}\mathrm{vec}\left(\tilde{A}\left(W_t\right)\right)\right\| &\leq K_{2,n}K_{1,n}\rho_M 2K_{3,3n}K_a && (32)\\
\sup_{\theta\in\Theta, t\in(-\infty,\infty)} \left\|\frac{\partial}{\partial a'}\mathrm{vec}\left(\frac{\partial}{\partial a'}\mathrm{vec}\left(\tilde{A}\left(W_t\right)\right)\right)\right\| &\leq K_{2,n}K_{1,n}\rho_M 2K_{3,3n}K_{1,3n} && (33)
\end{aligned}$$

In addition, in the following we will need bounds for the first and second derivatives of the products. For the first derivative we have:

$$\begin{aligned}
&\frac{\partial}{\partial b'}\mathrm{vec}\left(\left\{\prod_{l=r}^{s}\tilde{B}\left(W_{t-l}\right)\right\}\right)\\
= &\sum_{l=r}^{s}\left\{\prod_{m'=s-l+1}^{s}\tilde{B}\left(W_{t-m'}\right)\right\}' \otimes \left\{\prod_{m=r}^{s-l-1}\tilde{B}\left(W_{t-m}\right)\right\}\frac{\partial}{\partial b'}\mathrm{vec}\left(\tilde{B}\left(W_{t-(s-l)}\right)\right)
\end{aligned}$$

$$\begin{aligned}
&\sup_{\theta\in\Theta, t\in(-\infty,\infty)} \left\|\frac{\partial}{\partial b'}\mathrm{vec}\left(\left\{\prod_{l=r}^{s}\tilde{B}\left(W_{t-l}\right)\right\}\right)\right\|\\
\leq &\sum_{l=r}^{s}\left\{\sup_{\theta\in\Theta, t\in(-\infty,\infty)}\left\|\left\{\prod_{m'=s-l+1}^{s}\tilde{B}\left(W_{t-m'}\right)\right\}\right\| \sup_{\theta\in\Theta, t\in(-\infty,\infty)}\left\|\left\{\prod_{m=r}^{s-l-1}\tilde{B}\left(W_{t-m}\right)\right\}\right\| \times\right.\\
\times &\left.\sup_{\theta\in\Theta, t\in(-\infty,\infty)}\left\|\frac{\partial}{\partial b'}\mathrm{vec}\left(\tilde{B}\left(W_t\right)\right)\right\|\right\}\\
\leq &\sum_{l=r}^{s}\bar{\varrho}_B^{l}\bar{\varrho}_B^{s-l-r}\sup_{\theta\in\Theta, t\in(-\infty,\infty)}\left\|\frac{\partial}{\partial b'}\mathrm{vec}\left(\tilde{B}\left(W_t\right)\right)\right\|\\
\leq &\left(s-r+1\right)\bar{\varrho}_B^{s-r}\sup_{\theta\in\Theta, t\in(-\infty,\infty)}\left\|\frac{\partial}{\partial b'}\mathrm{vec}\left(\tilde{B}\left(W_t\right)\right)\right\| && (34)
\end{aligned}$$

For the second derivative we need the following intermediate result:

$$\frac{\partial}{\partial b'}\text{vec}\left(\left\{\prod_{m'=s-l+1}^{s}\tilde{B}\left(W_{t-m'}\right)\right\}'\otimes\left\{\prod_{m=r}^{s-l-1}\tilde{B}\left(W_{t-m}\right)\right\}\right) = \left[I_{n(n+1)/2}\otimes C_{n(n+1)/2}\otimes I_{n(n+1)/2}\right]\times$$
$$\times\left\{\frac{\partial}{\partial b'}\text{vec}\left(\left\{\prod_{m'=s-l+1}^{s}\tilde{B}\left(W_{t-m'}\right)\right\}'\right)\otimes\text{vec}\left(\left\{\prod_{m=r}^{s-l-1}\tilde{B}\left(W_{t-m}\right)\right\}\right)+\right.$$
$$\left.+\text{vec}\left(\left\{\prod_{m'=s-l+1}^{s}\tilde{B}\left(W_{t-m'}\right)\right\}'\right)\otimes\frac{\partial}{\partial b'}\text{vec}\left(\left\{\prod_{m=r}^{s-l-1}\tilde{B}\left(W_{t-m}\right)\right\}\right)\right\}$$

by which, we get the bound:

$$\sup_{\theta\in\Theta,t\in(-\infty,\infty)}\left\|\frac{\partial}{\partial b'}\text{vec}\left(\left\{\prod_{m'=s-l+1}^{s}\tilde{B}\left(W_{t-m'}\right)\right\}'\otimes\left\{\prod_{m=r}^{s-l-1}\tilde{B}\left(W_{t-m}\right)\right\}\right)\right\|$$
$$\leq K_{1,n(n+1)/2}\left(l+1\right)\bar{\varrho}_B^{l}\bar{\varrho}_B^{s-l-r}\sup_{\theta\in\Theta,t\in(-\infty,\infty)}\left\|\frac{\partial}{\partial b'}\text{vec}\left(\tilde{B}\left(W_t\right)\right)\right\|$$
$$+K_{1,n(n+1)/2}\left(s-l-r+1\right)\bar{\varrho}_B^{s-l-r}\bar{\varrho}_B^{l}\sup_{\theta\in\Theta,t\in(-\infty,\infty)}\left\|\frac{\partial}{\partial b'}\text{vec}\left(\tilde{B}\left(W_t\right)\right)\right\|$$
$$=K_{1,n(n+1)/2}\left(s-r+1\right)\bar{\varrho}_B^{s-r}\sup_{\theta\in\Theta,t\in(-\infty,\infty)}\left\|\frac{\partial}{\partial b'}\text{vec}\left(\tilde{B}\left(W_t\right)\right)\right\| \tag{35}$$

For the second derivative we have:

$$\frac{\partial}{\partial b'}\text{vec}\left(\frac{\partial}{\partial b'}\text{vec}\left(\left\{\prod_{l=r}^{s}\tilde{B}\left(W_{t-l}\right)\right\}\right)\right) =$$
$$=\sum_{l=r}^{s}\frac{\partial}{\partial b'}\text{vec}\left(\left\{\prod_{m'=s-l+1}^{s}\tilde{B}\left(W_{t-m'}\right)\right\}'\otimes\left\{\prod_{m=r}^{s-l-1}\tilde{B}\left(W_{t-m}\right)\right\}\frac{\partial}{\partial b'}\text{vec}\left(\tilde{B}\left(W_{t-(s-l)}\right)\right)\right)$$
$$=\sum_{l=r}^{s}\left[\left[\frac{\partial}{\partial b'}\text{vec}\left(\tilde{B}\left(W_{t-(s-l)}\right)\right)\right]'\otimes I_{n(n+1)/2}\right]\times$$
$$\times\frac{\partial}{\partial b'}\text{vec}\left(\left\{\prod_{m'=s-l+1}^{s}\tilde{B}\left(W_{t-m'}\right)\right\}'\otimes\left\{\prod_{m=r}^{s-l-1}\tilde{B}\left(W_{t-m}\right)\right\}\right)$$
$$+\sum_{l=r}^{s}\left\{I_{3n}\otimes\left\{\left\{\prod_{m'=s-l+1}^{s}\tilde{B}\left(W_{t-m'}\right)\right\}'\otimes\left\{\prod_{m=r}^{s-l-1}\tilde{B}\left(W_{t-m}\right)\right\}\right\}\right\}\times$$
$$\times\frac{\partial}{\partial b'}\text{vec}\left(\frac{\partial}{\partial b'}\text{vec}\left(\tilde{B}\left(W_{t-(s-l)}\right)\right)\right)$$

and finally

$$\begin{aligned}
& \sup_{\theta\in\Theta, t\in(-\infty,\infty)} \left\| \frac{\partial}{\partial b'} \mathrm{vec}\left(\frac{\partial}{\partial b'} \mathrm{vec}\left(\left\{ \prod_{l=r}^{s} \tilde{B}\left(W_{t-l}\right) \right\} \right) \right) \right\| \\
\leq \; & \left(s-r+1\right)^2 \bar{\varrho}_B^{s-r} K_{1,n(n+1)/2} \sup_{\theta\in\Theta, t\in(-\infty,\infty)} \left\| \frac{\partial}{\partial b'} \mathrm{vec}\left(\tilde{B}\left(W_t\right) \right) \right\|^2 \\
+ \; & \left(s-r+1\right)^2 \bar{\varrho}_B^{s-r} \sup_{\theta\in\Theta, t\in(-\infty,\infty)} \left\| \frac{\partial}{\partial b'} \mathrm{vec}\left(\frac{\partial}{\partial b'} \mathrm{vec}\left(\tilde{B}\left(W_t\right) \right) \right) \right\| \qquad (36)
\end{aligned}$$

For part (i)

$$\begin{aligned}
\xi_t &= \tilde{C} + \tilde{A}\left(W_t\right) y_{t-1} + \sum_{\tau=1}^{\infty} \left\{ \prod_{l=1}^{\tau-1} \tilde{B}\left(W_{t-l}\right) \right\} \left(\tilde{C} + \tilde{A}\left(W_{t-\tau}\right) X_{t-\tau-1} \right) \\
\frac{\partial \xi_t}{\partial a'} &= X'_{t-1} \otimes I_{n(n+1)/2} \frac{\partial \mathrm{vec}\left(\tilde{A}\left(W_t\right) \right)}{\partial a'} \\
&+ \sum_{\tau=1}^{\infty} \left\{ \prod_{l=1}^{\tau-1} \tilde{B}\left(W_{t-l}\right) \right\} \left(\left(X'_{t-\tau-1} \otimes I_{n(n+1)/2} \right) \frac{\partial \mathrm{vec}\left(\tilde{A}\left(W_{t-\tau}\right) \right)}{\partial a'} \right) \qquad (37) \\
\frac{\partial \xi_t}{\partial b'} &= \sum_{\tau=1}^{\infty} \left[\left(\tilde{C} + \tilde{A}\left(W_{t-\tau}\right) X_{t-\tau-1} \right)' \otimes I_{n(n+1)/2} \right] \frac{\partial}{\partial b'} \mathrm{vec}\left(\left\{ \prod_{l=1}^{\tau-1} \tilde{B}\left(W_{t-l}\right) \right\} \right) \qquad (38)
\end{aligned}$$

By using equation (37), equation (38), 32,29 and 34 and applying the Hölder and Minkowski inequalities, we get

$$\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\Theta} \left\| \frac{\partial \tilde{\xi}_t}{\partial a'} \right\|^3 \right] &\leq \left\{ \sum_{\tau=0}^{\infty} \bar{\varrho}_B^{\tau-1} \left[\mathbb{E}\left[\sup_{\theta\in\Theta} \left\| \left(\left(X'_{t-\tau-1} \otimes I_{n(n+1)/2} \right) \frac{\partial \mathrm{vec}\left(\tilde{A}\left(W_{t-\tau}\right) \right)}{\partial a'} \right) \right\|^3 \right]^{1/3} \right\}^3 \\
&\leq \left\{ \sum_{\tau=0}^{\infty} \bar{\varrho}_B^{\tau-1} \left[\mathbb{E}\left[\left\| u_t \right\|^6 \right]^{1/6} K_{2,n} K_{1,n} \rho_M 2 K_{3,3n} K_a \right] \right\}^3 < \infty \\
\mathbb{E}\left[\sup_{\theta\in\Theta} \left\| \frac{\partial \tilde{\xi}_t}{\partial b'} \right\|^3 \right]. &\leq \left\{ \sum_{\tau=0}^{\infty} \left(\tau-1\right) \bar{\varrho}_B^{\tau-2} \left[K_c^2 + \mathbb{E}\left[\left\| u_t \right\|^6 \right]^{1/6} K_{2,n} K_{1,n} \rho_M 2 K_{3,3n} K_b \right] \right\}^3 < \infty
\end{aligned}$$

For part (ii)

$$
\begin{aligned}
\frac{\partial^2 \xi_t}{\partial a' \partial a'} &= \left(X'_{t-1} \otimes I_{n(n+1)/2}\right) \frac{\partial}{\partial a'} \left(\mathrm{vec} \left(\frac{\partial \mathrm{vec}\left(\tilde{A}\left(W_t\right)\right)}{\partial a'} \right) \right) \\
&+ \sum_{\tau=1}^{\infty} \left\{ \prod_{l=1}^{\tau-1} \tilde{B}\left(W_{t-l}\right) \right\} \left(\left(X'_{t-\tau-1} \otimes I_{n(n+1)/2}\right) \frac{\partial}{\partial a'} \left(\mathrm{vec} \left(\frac{\partial \mathrm{vec}\left(\tilde{A}\left(W_t\right)\right)}{\partial a'} \right) \right) \right) \quad (39)
\end{aligned}
$$

$$
\begin{aligned}
\frac{\partial^2 \xi_t}{\partial a' \partial b'} &= \sum_{\tau=1}^{\infty} \left[\left(\left(X'_{t-\tau-1} \otimes I_{n(n+1)/2}\right) \frac{\partial \mathrm{vec}\left(\tilde{A}\left(W_{t-\tau}\right)\right)}{\partial a'} \right)' \otimes I_{n(n+1)} \right] \times \\
&\times \left\{ \sum_{j=1}^{\tau-1} \left\{ \prod_{l=1}^{\tau-1-j} \tilde{B}\left(W_{t-l}\right)' \right\} \otimes \left\{ \prod_{l=\tau-1-(j-1)}^{\tau-1} \tilde{B}\left(W_{t-l}\right) \right\} \frac{\partial}{\partial b'} \mathrm{vec}\left(\tilde{B}\left(W_{t-\tau+j}\right)\right) \right\} \quad (40)
\end{aligned}
$$

$$
\frac{\partial^2 \xi_t}{\partial b' \partial b'} = \sum_{\tau=1}^{\infty} \left[\left(\tilde{C} + \tilde{A}\left(W_{t-\tau}\right) X_{t-\tau-1}\right)' \otimes I_{n(n+1)/2} \right] \frac{\partial}{\partial b'} \mathrm{vec} \left(\frac{\partial}{\partial b'} \mathrm{vec} \left(\left\{ \prod_{l=r}^{s} \tilde{B}\left(W_{t-l}\right) \right\} \right) \right). \quad (41)
$$

Then, from equation (39),equation (40), equation (41), 33,30 and 36 and the Hölder and Minkowski inequalities, we have

$$
\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\Theta} \left\| \frac{\partial^2 \tilde{\xi}_t}{\partial a' \partial a'} \right\|^3 \right] &\leq \left\{ \sum_{\tau=0}^{\infty} \bar{\varrho}_B^{\tau-1} K_{2,n} K_{1,n} \rho_M 2 K_{3,3n} K_{1,3n} \left[\mathbb{E}\left[\|u_t\|^6 \right]^{1/6} \right] \right\}^3 < \infty \\
\mathbb{E}\left[\sup_{\theta\in\Theta} \left\| \frac{\partial^2 \tilde{\xi}_t}{\partial b' \partial a'} \right\|^3 \right] &\leq \left\{ \sum_{\tau=0}^{\infty} (\tau-1) \bar{\varrho}_B^{\tau-2} K_{2,n}^2 K_{1,n}^2 \rho_M^2 2 K_{3,3n}^2 K_a K_b \left[\mathbb{E}\left[\|u_t\|^6 \right]^{1/6} \right] \right\}^3 < \infty \\
\mathbb{E}\left[\sup_{\theta\in\Theta} \left\| \frac{\partial^2 \tilde{\xi}_t}{\partial b' \partial b'} \right\|^3 \right] &\leq \left\{ \left[K_c^2 + \mathbb{E}\left[\|u_t\|^6 \right]^{1/6} K_{2,n} K_{1,n} \rho_M 2 K_{3,3n} K_b \right] \times \right. \\
&\times \left[\sum_{\tau=0}^{\infty} (\tau-1)^2 \bar{\varrho}_B^{\tau-3} K_{2,n}^2 K_{1,n}^2 \rho_M^2 2 K_{3,3n}^2 K_b^2 \right. \\
&+ \left. \left. \sum_{\tau=0}^{\infty} (\tau-1)^2 \bar{\varrho}_B^{\tau-2} K_{2,n} K_{1,n} \rho_M 2 K_{3,3n} K_{1,3n} \right] \right\}^3 < \infty
\end{aligned}
$$

□

A.4 Proof of theorem 3 on page 18

Proof. The proof is analogous to the proof of Theorem 2 in Hafner and Preminger (2009) if we substitute in part (ii) of the proof of their Lemma 2 their bound of $\mathbb{E}\left[\sup_{\theta\in\Theta}\left\|\Sigma_t - \tilde{\Sigma}_t\right\|\right]$

(equation 17 in their paper) with our bound in Lemma 5 . □

A.5 Proof of theorem 4 on page 19

Proof. The proof is again almost identical to the proof of the equivalent Theorem 3 in Hafner and Preminger (2009) if we replace in part (i) and (ii) of their proof of Lemma 3 the bound on the derivatives of $\tilde{\xi}_t$ with respect to parameters (equations 28,29 and 33 in their paper) with the bound we derive in Lemma 6. □

Table 1: Decomposition of $[\Sigma_t]_{i,j}$

	shock response (ARCH)	persistence (GARCH)
Constant	$[CC']_{i,j}$	
direct	$v^0_{i,t-1}v^0_{j,t-1}$	$\left[\Omega^{0,0}_{t-1}\right]_{i,j}$
indirect	$v^1_{L,i,t-1}v^1_{L,j,t-1} + v^1_{R,i,t-1}v^1_{R,j,t-1}$ $+v^1_{L,i,t-1}v^1_{R,j,t-1} + v^1_{R,i,t-1}v^1_{L,j,t-1}$	$\left[\Omega^{1,1}_{L,L,t-1}\right]_{i,j} + \left[\Omega^{1,1}_{R,R,t-1}\right]_{i,j}$ $+\left[\Omega^{1,1}_{L,R,t-1}\right]_{i,j} + \left[\Omega^{1,1}_{R,L,t-1}\right]_{i,j}$
mixed	$v^1_{L,i,t-1}v^0_{j,t-1} + v^0_{i,t-1}v^1_{L,j,t-1}$ $+v^1_{R,i,t-1}v^0_{j,t-1} + v^0_{i,t-1}v^1_{R,j,t-1}$	$\left[\Omega^{1,0}_{L,t-1}\right]_{i,j} + \left[\Omega^{0,1}_{L,t-1}\right]_{i,j}$ $+\left[\Omega^{1,0}_{R,t-1}\right]_{i,j} + \left[\Omega^{0,1}_{R,t-1}\right]_{i,j}$

Table 2: Specification tests for the Spatial-BEKK model on Daily Changes in the Five-Year EMU Sovereign CDS spreads from 9/10/2008 to 30/12/2016.

Joint Significance of Network Parameters				
	LR	Wald	Wald L	Wald R
Stat	268	176	97	55
P-value	0.0000	0.0000	0.0000	0.0000
	Wald $a_{1,L}$	Wald $a_{1,R}$	Wald $b_{1,L}$	Wald $b_{1,R}$
Stat	38	27	80	32
P-value	0.0000	0.0002	0.0000	0.0000

Table 3: Estimated Relevant Parameters of SB-BEKK on Daily Changes in the Five-Year EMU Sovereign CDS spreads from 9/10/2008 to 30/12/2016. Standard deviation in parenthesis. * parameters significant at the 10% level. ** parameters significant at the 5% level. *** parameters significant at the 1% level

	$a_{1,L}$	$a_{1,R}$	$b_{1,L}$	$b_{1,R}$	a_0	b_0
DE	-0.056(0.11)	-0.022(0.10)	-0.088(0.03)***	-0.014(0.03)	0.258(0.03)***	0.965(0.01)***
IT	0.089(0.09)	0.849(0.27)***	0.001(0.03)	-0.058(0.07)	0.192(0.02)***	0.972(0.00)***
FR	-0.636(0.19)***	0.549(0.22)**	0.209(0.05)***	-0.292(0.10)***	0.376(0.03)***	0.887(0.01)***
IE	-0.388(0.11)***	-1.719(0.53)***	0.149(0.03)***	0.757(0.16)***	0.352(0.04)***	0.929(0.01)***
ES	0.618(0.18)***	-0.568(0.22)***	-0.138(0.04)***	0.076(0.04)*	0.107(0.02)***	0.987(0.00)***
PT	0.206(0.12)*	-0.742(0.10)***	0.086(0.04)**	0.252(0.01)***	0.290(0.03)***	0.912(0.01)***

Table 4: The percentage of cumulative variance decomposition was obtained from data on daily changes in the five-year EMU sovereign CDS spreads from 9/10/2008 to 30/12/2016.

	Constant	Mixed	Indirect	Direct
VC_{DE} %	0.0170	-0.0104	0.0021	0.9913
VC_{IT} %	0.0170	-0.0034	0.0019	0.9845
VC_{FR} %	0.0563	0.0137	0.0104	0.9196
VC_{IE} %	0.0167	0.0063	0.0037	0.9732
VC_{ES} %	0.0209	-0.0282	0.0192	0.9881
VC_{PT} %	0.0344	0.0436	0.0059	0.9161

Table 5: The percentage of cumulative marginal spillover contribution (MSC) was obtained from data on daily changes in the five-year EMU sovereign CDS spreads from 9/10/2008 to 30/12/2016.

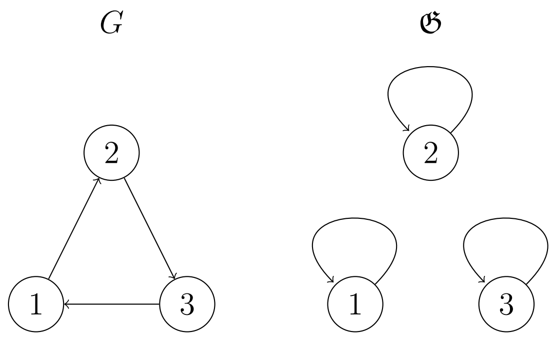
	Total	Costant	Mixed	Indirect	Direct
MSC_{DE} %	0.0893	0.0015	0.0011	0.0000	0.0866
MSC_{IT} %	0.0984	0.0022	0.0063	0.0005	0.0894
MSC_{FR} %	0.1442	0.0058	-0.0011	0.0010	0.1385
MSC_{IE} %	0.1413	0.0023	0.0071	-0.0005	0.1323
MSC_{ES} %	0.2566	0.0059	0.0042	0.0002	0.2463
MSC_{PT} %	0.2692	0.0037	0.0019	-0.0002	0.2638

Table 6: Parameter-constancy test for the different optimal networks

	No Constraints	Redistribution Constraint	Frobenius Constraint
Wald Stat	27.1	8.2	7.4
P-value	0.3012	0.9988	0.9995

Table 7: Investment Needed to Reach Target Exposures (Billions of USD), obtained from daily changes in five-year EMU sovereign CDS spreads.

	Optimal Network with Frobenius Constraint (Delta wrt true) Billions of USD					
Location	DE	IT	FR	IE	ES	PT
Reporting						
DE		-41.2	32.6	-5.6	41.5	109.7
IT	8.1		-6.5	-7.7	-11.4	15.3
FR	44.9	-121.9		-10.8	9.0	85.3
IE	-0.8	-1.3	-0.7		-1.9	2.4
ES	-10.9	-56.3	-19.8	-2.8		24.1
PT	-0.7	-3.6	-1.7	-0.4	-0.9	



$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{pmatrix} \qquad \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Figure 1: The directed triangle G has one strongly connected component but induces a $\mathfrak{G}$ with 3 components

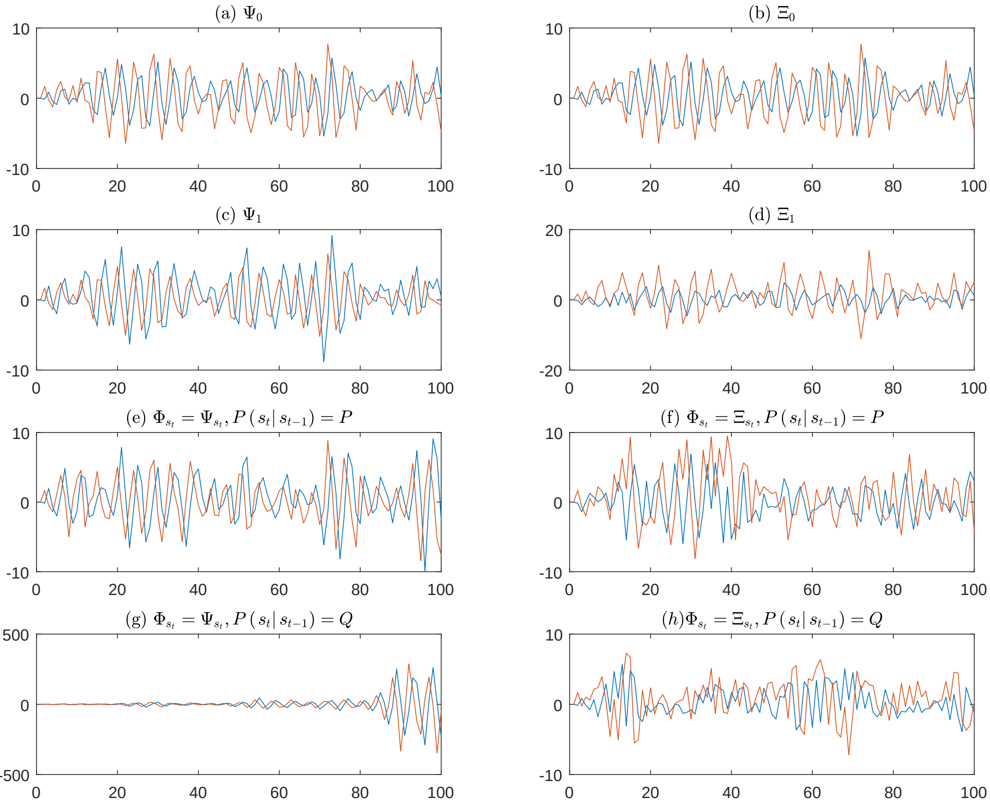


Figure 2: Simulated paths of : (a) bivariate VAR(1) with autocorrealtion matix Ψ_0; (c) bivariate VAR(1) with autocorrealtion matix Ξ_0; (d) bivariate VAR(1) with autocorrealtion matix Ψ_1; (d) bivariate VAR(1) with autocorrealtion matix Ξ_0; (e) Markov Switching bivariate VAR(1) with $\Phi_{s_t} = \Psi_{s_t}$ and transition matrix P; (f) Markov Switching bivariate VAR(1) with $\Phi_{s_t} = \Xi_{s_t}$ and transition matrix P; (g) Markov Switching bivariate VAR(1) with $\Phi_{s_t} = \Psi_{s_t}$ and transition matrix Q; (h) Markov Switching bivariate VAR(1) with $\Phi_{s_t} = \Xi_{s_t}$ and transition matrix Q.

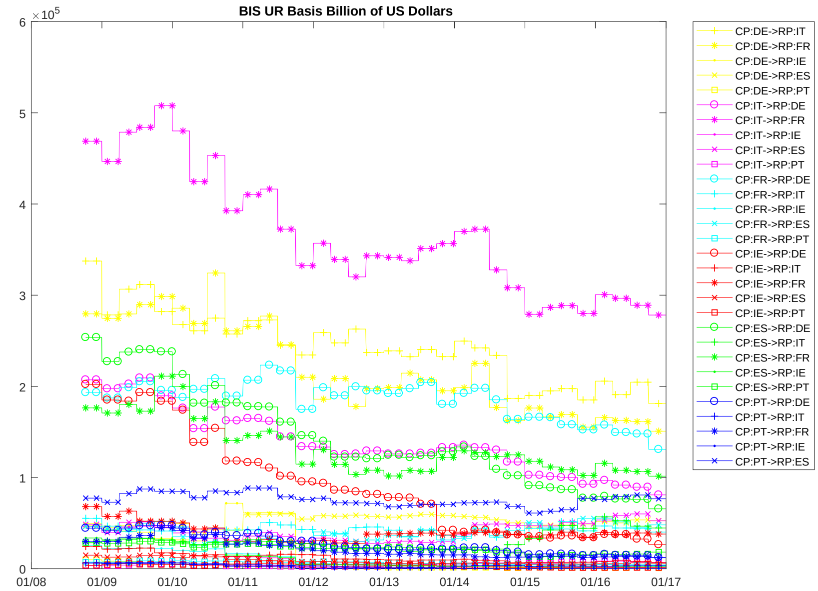


Figure 3: BIS claims in billions of US Dollars by Counterparty (CP) and Reporting country(RP)

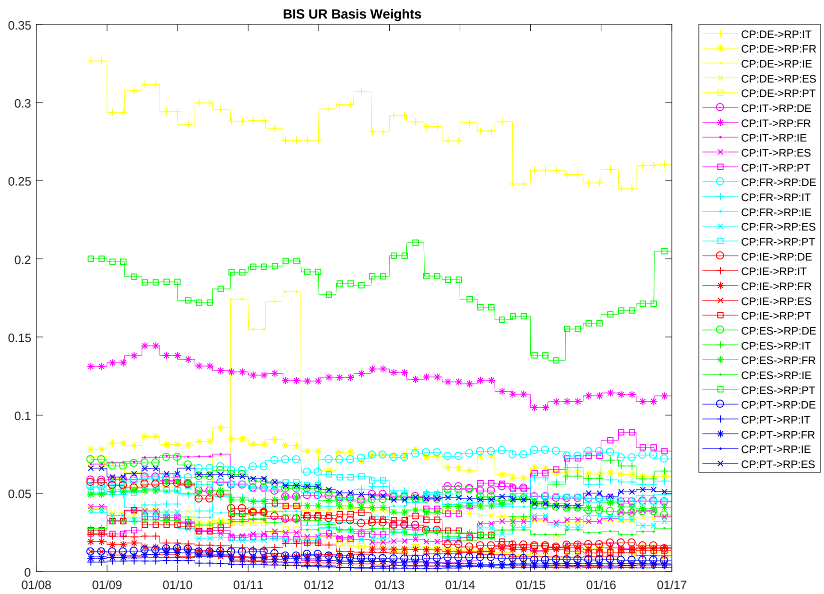


Figure 4: BIS claims normalized by the worldwide amount declared by reporting country, by Counterparty (CP) and Reporting country (RP)

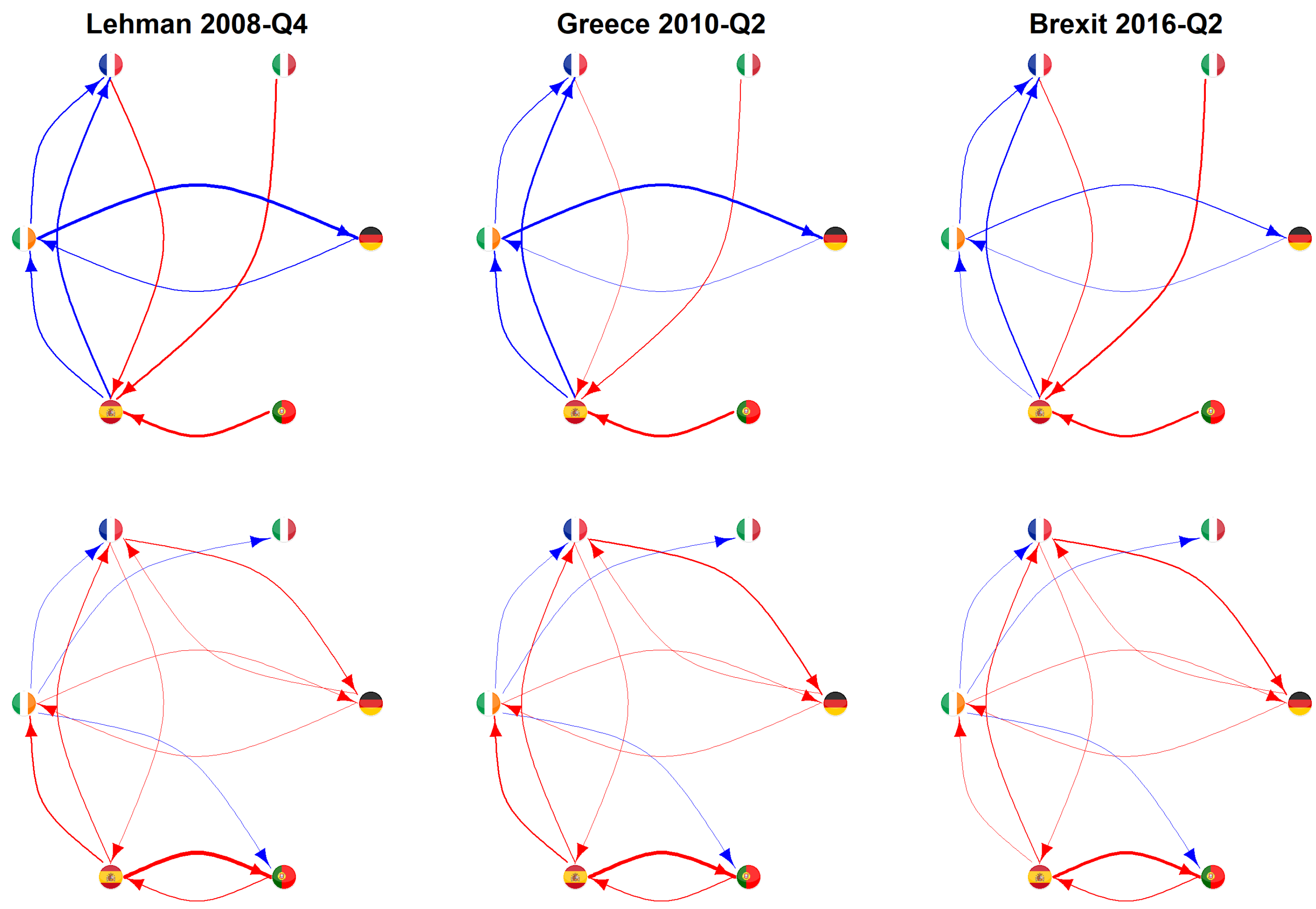


Figure 5: Selected graphical representations of the shock-response networks $\mathcal{W}_t^A$ in the first row and the persistence networks $\mathcal{W}_t^B$ in the second row were all obtained from daily changes in the five-year EMU sovereign CDS spreads from 9/10/2008 to 30/12/2016. The arrows go from counterparty to reporting, and positive weights are in red while negative are in blue. Line thickness is proportional to coefficent size. We only report edges that are significant at the 5% level, taking into consideration a familywise Bonferroni correction.

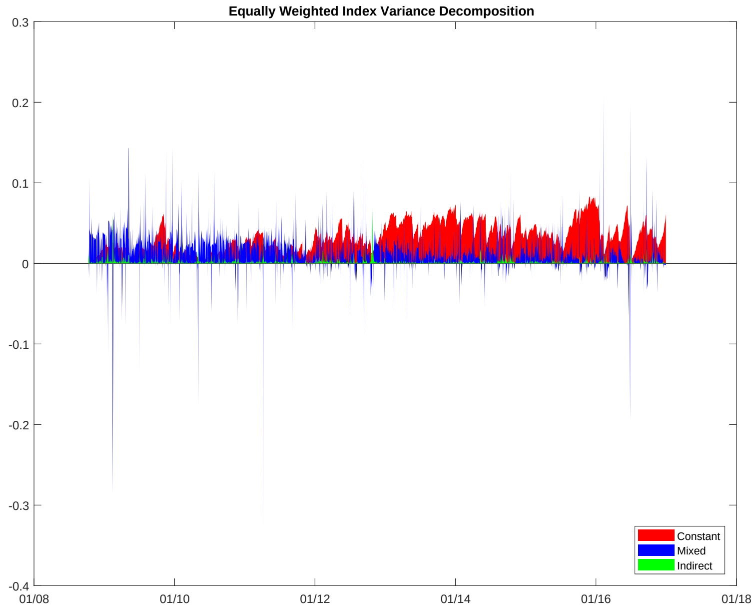


Figure 6: The relative variance decomposition of the equally weighted index was obtained from data on daily changes in the five-year EMU sovereign CDS spreads from 9/10/2008 to 30/12/2016, with the direct contribution omitted.

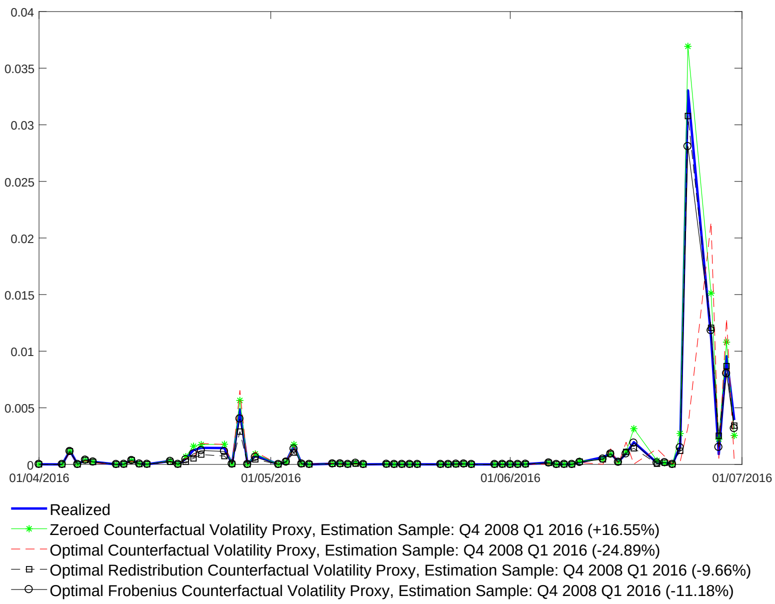


Figure 7: The average covariance proxy during the Brexit referendum of Q2 2016 was obtained from data on daily changes in five-year EMU sovereign CDS spreads. The percentage of cumulative proxy change with regard to the realized one is reported in parenthesis