

Mauro Bernardi – Michele Costola

High-Dimensional Sparse Financial Networks through a Regularised Regression Model

SAFE Working Paper No. 244

SAFE | Sustainable Architecture for Finance in Europe

A cooperation of the Center for Financial Studies and Goethe University Frankfurt

House of Finance | Goethe University
Theodor-W.-Adorno-Platz 3 | 60323 Frankfurt am Main

Tel. +49 69 798 30080 | Fax +49 69 798 33910
info@safe-frankfurt.de | www.safe-frankfurt.de

Non-Technical Summary

Given the recent financial crises, the analysis of the financial system and its interconnectedness has become increasingly relevant for policy authorities to monitor financial stability. Literature showed that direct interconnections among financial institutions can represent one of the main drivers of systemic risk. In this respect, network analysis proved to be an efficient approach to measure connectedness in the financial system.

We propose a shrinkage and selection methodology specifically designed for network inference with high-dimensional data and present a linear regression model with Spike-and-Slab prior on the coefficients. The approach is then routinely adapted to estimate a high-dimensional VAR equation-by-equation, where the dimension of regressors is larger than the number of observations, preventing the estimation via the joint likelihood. The proposed model can be interpreted as a VAR model with a diagonal variance-covariance matrix where shocks are instantaneously uncorrelated and the number of covariates is strictly larger than the number of observations. We aim to improve the identification of the relationship (linkages) among the financial firms by including a very large number of institutions in each equation by maintaining de facto the flexibility of the univariate framework. Moreover, the model accounts for two sets of covariates. The first set contains the predetermined variables which cannot be penalized: these variables are the autoregressive component for the persistence of the given financial institution (the dependent variable) and a set of common factors which model the cross-section correlation in the residuals to ensure independence. The second set of variables contains all the (lagged) financial institutions in the system that can be included at a given probability. The inference of the network is straightforward since the financial linkages are derived by the (posterior) inclusion probability of a given institution. Consequently, the weighted directed network is obtained through the adjacency matrix which is built “row by row”, that is equation by equation, using for each firm the posterior inclusion probabilities of all the others in the financial system.

In the empirical application, we consider the weekly closing price series of 1248 world financial firms (banks, insurances, brokers, and other financial services), both active and dead from 29 December 2000 to 6 October 2017. We estimate the dynamic financial network using a rolling window approach and then focus on the connectedness over time by analyzing the network out-degrees. Usually, the presence of heterogeneity in the financial linkages results in scale-free networks having the well-known heavy-tailed property of the degree distribution. This indicates a

majority (minority) group of firms having a low (high) number of connections. In this respect, we make use of the Generalized Pareto distribution to model the shape of the out-degrees distribution. The variation over time of the shape parameter of the distribution can represent a signal of structural changes in network connectedness detecting whether there is a mutation on the number of network hubs (i.e., nodes with higher degrees). Interestingly, the dynamic of this hubs indicator exhibits the typical behavior of financial stress indicators and promptly reacts during significant economic and financial events. Given its persistence over time, we select the MSCI World Financials Index as a proxy for the considered global market and show, after controlling for volatility and other robustness checks, that it represents a significant predictor of market returns at the first lag (one week) and the fourth lag (one month). These findings demonstrate that the proposed approach can be successfully applied to the estimation of large financial networks and can be used to construct real-time financial stress indicators.

High-Dimensional Sparse Financial Networks through a Regularised Regression Model

Mauro Bernardi^{*1} and Michele Costola^{†2}

¹Department of Statistical Sciences, University of Padova and Istituto per le Applicazioni del Calcolo “Mauro Picone” - CNR, Roma, Italy

²Research Center SAFE, House of Finance, Goethe University Frankfurt, Germany

February 12, 2019

Abstract

We propose a shrinkage and selection methodology specifically designed for network inference using high-dimensional data through a regularised linear regression model with Spike-and-Slab prior on the parameters. The approach extends the case where the error terms are heteroscedastic, by adding an ARCH-type equation through an approximate Expectation-Maximisation algorithm. The proposed model accounts for two sets of covariates. The first set contains predetermined variables which are not penalised in the model (i.e., the autoregressive component and common factors) while the second set of variables contains all the (lagged) financial institutions in the system, included with a given probability. The financial linkages are expressed in terms of inclusion probabilities resulting in a weighted directed network where the adjacency matrix is built “row by row”. In the empirical application, we estimate the network over time using a rolling window approach on 1248 world financial firms (banks, insurances, brokers and other financial services) both active and dead from 29 December 2000 to 6 October 2017 at a weekly frequency. Findings show that over time the shape of the out-degree distribution exhibits the typical behaviour of financial stress indicators and represents a significant predictor of market returns at the first lag (one week) and the fourth lag (one month).

KEYWORDS: VAR estimation, Financial Networks, Bayesian inference, Sparsity, Spike-and-Slab prior, Stochastic Search Variable Selection, Expectation-Maximisation.

^{*}Corresponding author. Via C. Battisti, 241, 35121 Padua, Italy. e-mail: mauro.bernardi@unipd.it, Phone.: +39.049.8274165.

[†]Theodor-W.-Adorno-Platz 3, 60323 Frankfurt am Main, Germany. e-mail: costola@safe.uni-frankfurt.de, Phone: +49 69 798 34505, Fax +49 69 798 30077.

1 Introduction

Given the recent financial crises, the analysis of the financial system and its interconnectedness has become increasingly relevant for policy authorities to monitor financial stability. Among others, [Bluhm and Krahnen \(2014\)](#) show that direct interconnections among financial institutions represents one of the main drivers of systemic risk. In this respect, network analysis proved to be an efficient approach to measure connectedness in the financial system. Scholars have introduced several theoretical models on financial networks ([Allen and Babus, 2009](#), [Elliott et al., 2014](#), [Acemoglu et al., 2015](#)) and inference approaches ([Billio et al., 2012](#), [Diebold and Yilmaz, 2014](#), [Ahelegbey et al., 2016](#), [Billio et al., 2016](#), [Demirer et al., 2017](#), [Korobilis and Yilmaz, 2018](#), [Billio et al., 2018](#), [Barigozzi and Brownlees, 2019](#), [Bianchi et al., 2019](#), [Billio et al., 2019](#)).

Regarding the recent literature, [Billio et al. \(2012\)](#) propose pairwise Granger causality tests to describe causal linkages among firms' returns. This approach is very flexible since it can be virtually applied to datasets of any dimension, but at the same time, the transmission mechanism (linkages) may be not correctly identified given that shocks are not jointly considered. As noted in [Acharya et al. \(2012\)](#), it is not possible to distinguish if one firm Granger causes another one or a third one is causing both. In this regard, multivariate models such as Vector Autoregressive (VAR) represent one of the best ways to describe structural shocks in a given system. For instance, [Ahelegbey et al. \(2016\)](#) show that VAR models provide a better representation of the linkages than Granger-causality tests. Nevertheless, the classical multivariate framework suffers of the well known curse of dimensionality which limits the range of applicability to relatively small cross-sectional sizes. To overcome this issue, shrinkage and selection methodologies have been proposed to deal with high-dimensional data. The first and the most popular is the least absolute shrinkage and selection operator (LASSO) introduced by [Tibshirani \(1996\)](#) which can simultaneously perform parameters estimation and selection in regression models. After that, further literature introduced alternative shrinkage methods such as the least angle regression (LARS) of [Efron et al. \(2004\)](#), the boosting of [Bühlmann \(2006\)](#), the adaptive LASSO of [Zou \(2006\)](#) and the group LASSO of [Yuan and Lin \(2006\)](#). In a multivariate context, sparse methods that automatically shrink to zero the off diagonal elements of the variance-covariance matrix are crucial because they reduce the number of estimated parameters. Several works related to sparse estimation of either the variance-covariance matrix or its inverse are available in literature. [Friedman et al. \(2008\)](#) propose a fast algorithm based on a coordinate-wise updating scheme in order to estimate a sparse graph using the least absolute shrinkage and selection operator (LASSO) ℓ_1 -penalty of [Tibshirani \(1996\)](#). [Meinshausen and Bühlmann \(2006\)](#) present a method for neighbourhood selection using the LASSO ℓ_1 -penalty as an alternative to covariance selection for Gaussian graphical models where the number of observations is less than the number

of variables. Gao and Massam (2015) estimate the variance–covariance matrix of symmetry–constrained Gaussian models using three different ℓ_1 –type penalty functions, i.e., the LASSO, the smoothly clipped absolute deviation (SCAD) of Fan and Li (2001) and the minimax concave penalty (MCP) of Zhang (2010). Bien and Tibshirani (2011) propose a penalised version of the log–likelihood function, using the LASSO penalty, in order to estimate a sparse variance–covariance matrix of a multivariate Gaussian distribution. Rothman et al. (2010) introduced the Multivariate Regression with Covariance Estimation (MRCE), a two–stage iterative procedure to achieve sparse estimation of both the regression parameters and the variance–covariance matrix within a multivariate regression framework that can be easily adapted to stochastic regressors in a VAR approach. From a Bayesian perspective, George and McCulloch (1993) proposed the Stochastic Search Variable Selection (SSVS) approach and a Gibbs sampler algorithm that enable to simulate draws from the posterior distribution of model parameters augmented by an indicator that includes the selected covariates. The SSVS method advocated by George and McCulloch (1993) has been gaining popularity as with Li and Zhang (2010) and Scheipl et al. (2012), although it suffers from a heavy computational burden.

Selection and shrinkage operators have been also adopted in the estimation of financial networks. For example, Demirer et al. (2017) apply the variance decomposition approach of Diebold and Yilmaz (2014) to estimate a global bank network which includes the world’s top 150 banks during the period from 2003 to 2014. More recently, Barigozzi and Brownlees (2019) propose the NETS algorithm based on LASSO and model a large VAR (90 US blue chips across different industries) where the autoregressive matrices and the inverse covariance matrix of the innovations are sparse.

Our paper aims to contribute to this stream of literature. We propose a shrinkage and selection methodology specifically designed for network inference with high–dimensional data and present a linear regression model with Spike–and–Slab prior on the coefficients. The methodology is then routinely adapted to estimate a high–dimensional VAR equation–by–equation, where the dimension of regressors is larger than the number of observations, preventing the estimation via the joint likelihood. Specifically, we consider a regularised linear regression model with the Spike–and–Slab prior on the coefficients (see, e.g., Mitchell and Beauchamp, 1988, George and McCulloch, 1993, 1997) which consists of a finite mixture of two Gaussian distributions having the same mean and different variances. The marginal inclusion probability of each regressor is obtained by averaging the posterior draws of the mixture indicator that refer to the component with higher variance. Therefore, a Markov chain Monte Carlo (MCMC) sampling algorithm should be implemented that efficiently simulates from the complete set of full conditional distributions. Unfortunately, simulating from the space of regressors augmented by the inclusion indicators is unsuitable where VAR parameters are routinely estimated equation–by–equation. Indeed, the large dimension of the state space augmented by the regressor indicators requires

high computational costs for approximating the posterior via MCMC for each equation estimation. To avoid the computational burden of MCMC methods we instead pursue the idea of Ročková and George (2014) to develop a fast and efficient Expectation–Maximisation (EM) algorithm (Dempster et al., 1977) in order to compute the Maximum–a–Posteriori (MaP) of a linear regression model with Spike–and–Slab prior. The resulting Stochastic Search Variable Selection Expectation–Maximisation (SSVS–EM) algorithm provides the MaP of the inclusion probability of each regressor, avoiding the computation of the whole posterior distribution. Ročková and George (2014) show that the SSVS–EM algorithm is an accurate deterministic approach with enormous computational savings, and has the potential to be a key player for variable selection problems. Here, we extend the SSVS–EM approach to the case where the error terms of the Gaussian linear regression model are heteroscedastic by adding an ARCH–type equation (Engle, 1982, Engle and Bollerslev, 1986) to model the dynamic evolution of the variance. Modelling the conditional variance is a relevant issue when dealing with dependent data that cannot be standardised before the application of the SSVS method. From an algorithmic perspective, including a dynamic variance results in a set additional parameters that enters nonlinearly the likelihood equation prevents an explicit analytical solution for the maximisation step of the EM algorithm. Therefore, we deliver an approximate–EM algorithm where the likelihood contribution of the ARCH parameters is approximated around the current values by a Gaussian distribution leading to optimal updates of the parameters. In a different context, an approximated EM algorithm has been proposed by Rizopoulos et al. (2009) to approximate the expectation step in longitudinal survival models.

The proposed model can be interpreted as a VAR model with a diagonal variance–covariance matrix where shocks are instantaneously uncorrelated and the number of covariates is strictly larger than the number of observations. We aim to improve the identification of the relationship (linkages) among the financial firms by including a very large number of institutions in each equation by maintaining de facto the flexibility of the univariate framework. Moreover, the model accounts for two sets of covariates. The first set contains the predetermined variables which cannot be penalised: these variables are the autoregressive component for the persistence of the given financial institution (the dependent variable) and a set of common factors which model the cross–section correlation in the residuals to ensure independence. The second set of variables contains all the (lagged) financial institutions in the system that can be included at a given probability. The inference of the network is straightforward since the financial linkages are derived by the (posterior) inclusion probability of a given institution. Consequently, the weighted directed network is obtained through the adjacency matrix which is built “row by row”, that is equation by equation, using for each firm the posterior inclusion probabilities of all the others in the financial system.

In the empirical application, we consider the weekly closing price series of 1248 world financial firms (banks, insurances, brokers and other financial services), both active and dead from 29 December 2000 to 6 October 2017. We estimate the dynamic financial network using a rolling window approach and then focus on the connectedness over time by analysing the network out-degrees. Usually, the presence of heterogeneity in the financial linkages results in scale-free networks having the well-known heavy-tailed property of the degree distribution. This indicates a majority (minority) group of firms having a low (high) number of connections (see Barabási and Albert, 1999, Barabási and Bonabeau, 2003, Boss et al., 2004). In this respect, we make use of the Generalized Pareto distribution to model the shape of the out-degrees distribution. The variation over time of the shape parameter of the distribution can represent a signal of structural changes in network connectedness detecting whether there is a mutation on the number of network hubs (i.e., nodes with higher degrees). Interestingly, the dynamic of this hubs indicator exhibits the typical behaviour of financial stress indicators and promptly reacts during significant economic and financial events. Given its persistence over time, we select the MSCI World Financials Index as a proxy for the considered global market and show, after controlling for volatility and other robustness checks, that it represents a significant predictor of market returns at the first lag (one week) and the fourth lag (one month). These findings demonstrate that the proposed approach can be successfully applied to the estimation of large financial networks and can be used to construct real time financial stress indicators.

The remaining of the paper is structured as follows. Section 2 introduces and formalises the main problem we address in this paper, namely the identification and estimation of large dimensional networks using dynamic regression models. Section 3 introduces the sparse heteroscedastic linear regression approach used to estimate the network and the SSVS-EM algorithm that uses a Spike-and-Slab prior to select the relevant coefficients. Section 4 illustrates the empirical analysis and our major findings, and Section 5 concludes the paper.

2 General framework for network estimation

In this Section, we introduce the general framework we propose for the network inference using the Vector Autoregressive (VAR) approach. The distinguishing feature of the proposed model which is relevant for network estimation, is that it deals with two different sets of covariates. The first set contains the predetermined variables which cannot be excluded from the regression such as the autoregressive component of the considered financial firms, that is the dependent variable in each equation of the system, and a set of common factors which account for the cross-section correlation of the residuals which lead to uncorrelated error terms. The second set of variables contains all the (lagged) returns of all the other financial firms in the system

that can be included in the regression model with a given probability. To this end, we develop a framework for the equation-by-equation estimation of the VAR parameters using a Bayesian sparse heteroscedastic linear regression method. An additional latent indicator having a Spike-and-Slab prior distribution is included to probabilistically assess the relevance of each institution in the network. The next Section describes the VAR model specification and its link with the proposed equation-by-equation estimation approach.

2.1 VAR model specification

Consider the random vector $\mathbf{y}_t = (y_{1,t}, y_{2,t}, \dots, y_{N,t})^\top \in \mathbb{R}^N$ collecting the cross-section of log-returns $y_{i,t} \in \mathbb{R}$, $i = 1, 2, \dots, N$ of N financial institutions at times $t = 1, 2, \dots, T$. We consider the following VAR model which expresses the cross-section returns of all the institutions at time t , \mathbf{y}_t , as a function of the lagged returns:

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{j=1}^k \boldsymbol{\Phi}_j \mathbf{y}_{t-j} + \boldsymbol{\eta}_t, \quad t = 2, 3, \dots, T, \quad (1)$$

where $\boldsymbol{\mu} \in \mathbb{R}^N$ is a vector of constant terms, $\boldsymbol{\Phi}_j$ is the $(N \times N)$ autoregressive matrix of loading parameters at the j -th lag and $\boldsymbol{\eta}_t \sim \mathcal{N}(0, \boldsymbol{\Sigma}_\eta)$ is the vector of error terms that are assumed to follow a multivariate Gaussian distribution with variance-covariance matrix $\boldsymbol{\Sigma}_\eta$. As explained in the Introduction and below in Section 4, the cross-sectional dimension of the vector in our application is very large, i.e., $N \gg 100$ preventing the estimation of the VAR parameters $(\boldsymbol{\Phi}_j, j = 1, 2, \dots, k, \boldsymbol{\Sigma}_\eta)$ using likelihood-based procedures. Nevertheless the VAR specification in equation (1) can be viewed as a seemingly unrelated regressions (SUR) model where each equation shares the same set of covariates. Therefore, the autoregressive matrices $\boldsymbol{\Phi}_j$ for $j = 1, 2, \dots, p$ in equation (1) can be consistently estimated using an equation-by-equation strategy through ordinary least squares. Unfortunately, the previous approach assumes that the off-diagonal elements in the covariance matrix of the error term $\boldsymbol{\eta}_t$ are zero, thereby leading to a biased parameters estimate. To account for cross-sectional correlation in the residuals while retaining a tractable equation-by-equation approach, we model $\boldsymbol{\eta}_t$ through an $f \times 1$ vector of common factors \mathbf{f}_t . One of the advantages with the inclusion of a vector of common factors in equation (1) relies on the ability to distinguish the systematic component measured by the factors and the idiosyncratic one related to the financial firm. Furthermore, the vector of common factors can be either observable or latent leading to fixed-effects or random-effects regression models, respectively. In this paper, we focus on the former case for two reasons. First, observed factors have a clear economic interpretation. Second, given the high dimension of the regressors space, observable factors are preferable in our case since they lead to a more

parsimonious representation of the common components. Accordingly, the correlation structure of $\boldsymbol{\eta}_t$ in equation (1) is accounted for by including in the VAR specification a set of common factors \mathbf{f}_t with the following specification:

$$\boldsymbol{\eta}_t = \mathbf{F}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \quad t = 2, 3, \dots, T, \quad (2)$$

where \mathbf{F} is a matrix of dimension $(N \times f)$ of factor loadings and $\boldsymbol{\epsilon}_t \sim \mathbf{N}(0, \boldsymbol{\Sigma}_\epsilon)$ contains the idiosyncratic components of $\boldsymbol{\eta}_t$ which are mutually uncorrelated such that $\boldsymbol{\Sigma}_\epsilon = \text{diag}\{\sigma_{1,1}^\epsilon, \sigma_{2,2}^\epsilon, \dots, \sigma_{N,N}^\epsilon\}$ is a diagonal matrix. Therefore, the following factor VAR model is obtained through equations (1) and (2):

$$\mathbf{y}_t = \boldsymbol{\mu} + \sum_{j=1}^k \boldsymbol{\Phi}_j \mathbf{y}_{t-j} + \mathbf{F}\mathbf{f}_t + \boldsymbol{\epsilon}_t, \quad t = 2, 3, \dots, T. \quad (3)$$

Without loss of generality, the VAR model in equation (3) can be linearly estimated equation-by-equation since the error terms $\boldsymbol{\epsilon}_t$ are uncorrelated by construction, while maintaining a certain degree of heteroscedasticity that characterises financial returns (see, e.g., [Francq and Zakoïan, 2010](#)). In this respect, the estimation equation-by-equation approach detailed below also accounts for the time-varying nature of the second moment of the returns' distribution that would highly influence the efficiency of the parameters estimate. Specifically, we assume an ARCH-type process ([Engle, 1982](#), [Bollerslev, 1986](#)) for the variance of the innovations $\epsilon_{i,t}$ in each regression equation. To highlight the estimation strategy, the proposed VAR model in equation (3) is written in terms of the i -th equation, that is, equation-by-equation:

$$y_{i,t} = \mu_i + \sum_{l=1}^k \varrho_{i,l} y_{i,t-l} + \mathbf{x}_{i,t}^\top \boldsymbol{\beta}_i + \mathbf{F}_i \mathbf{f}_t + h_{i,t}^{1/2} \varepsilon_{i,t}, \quad \varepsilon_{i,t} \stackrel{\text{i.i.d.}}{\sim} \mathbf{N}(0, 1), \quad (4)$$

$$h_{i,t} = \gamma_{i,h} + \alpha_{i,h} \varepsilon_{i,t-1}^2 + \beta_h h_{i,t-1}, \quad i = 1, 2, \dots, N, \quad t = 2, 3, \dots, T, \quad (5)$$

where μ_i is the intercept, $y_{i,t-l}$ is the i -th autoregressive component, $\mathbf{x}_{i,t} = (\mathbf{y}_{-i,t-1}^\top, \mathbf{y}_{-i,t-2}^\top, \dots, \mathbf{y}_{-i,t-k}^\top)^\top$ is the $k(N-1) \times 1$ vector of the lagged endogenous variables without the i -th element, and \mathbf{F}_i is the i -th row vector of the factor loadings matrix. Equation (5) specifies a weekly stationary GARCH(1,1) process of [Bollerslev \(1986\)](#) for the conditional variance terms $h_{i,t} \in \mathbb{R}^+$ with parameters $\gamma_h > 0$, $\alpha_h \geq 0$ and $\beta_h \geq 0$ and $\alpha_h + \beta_h < 1$. We collect the GARCH parameters in the vector $\boldsymbol{\psi} = (\gamma_h, \alpha_h, \beta_h)$. The main purpose of including the GARCH dynamics for the conditional variance in the equation-by-equation approach is to increase the statistical efficiency of the estimated regression effects $(\varrho_{i,1}, \varrho_{i,2}, \dots, \varrho_{i,k}, \boldsymbol{\beta}_i^\top, \mathbf{F}_i)$ while retaining a simple,

effective and computationally tractable model. During the years, financial econometrics literature introduced several GARCH specifications to account for financial stylised facts such as the asymmetric response of volatility to past shocks (see survey examples: [Hentschel et al., 1995](#), [Engle, 2001](#), [Francq and Zakoïan, 2010](#)). In several cases, those models represent an improvement in volatility forecasting (i.e., [Hansen and Lunde, 2005](#), [Awartani and Corradi, 2005](#)) which clearly is not the objective of this paper. Therefore, although more sophisticated models are appealing, we opt for the simpler yet parsimonious standard GARCH model that highly reduces the computational time. Anyhow, other GARCH specifications can be easily adapted to the estimation methodology proposed below.

The heteroscedastic Gaussian linear regression framework defined in equations (4)–(5) is a statistical tool for performing estimation of the parameters of the VAR model. Despite its usefulness for solving the curse of the dimensionality problem for high-dimensional VAR estimation, the equation-by-equation regression framework requires repeatedly solving the estimation issue which increases the computational burden. Moreover, usually the number of observations T is smaller than the number of covariates. Previous considerations, as well as the need of a probabilistic network structure that links financial institutions as vertexes, motivate the use of the Bayesian approach to variable selection. Indeed, the Bayesian approach here provides a valid inferential tool for performing both parameters estimation and model selection where the latter task aims to properly define the structure of the network. This approach requires the specification of a prior distribution for all the parameters. We deal with prior formulation in Section 3.1. However, for what follows, the prior specification for the regression parameters β_i in equation (4) is particularly crucial. For those parameters we introduce a Spike-and-Slab prior ([Mitchell and Beauchamp, 1988](#), [George and McCulloch, 1993, 1997](#)), which consists of a finite mixture of two Gaussian distributions with different variances. The Spike-and-Slab approach automatically delivers the marginal inclusion probability of each regressor thereby providing a natural framework for building the network structure which is described in the next Section (2.2). Bayesian methods usually involve computationally intensive simulation methods to numerically approximate the posterior distribution. To avoid the computational burden of MCMC methods, we pursue the idea of [Ročková and George \(2014\)](#) for developing a fast and efficient Expectation–Maximisation (EM) algorithm ([Dempster et al., 1977](#)) to compute the Maximum–a–Posteriori (MaP). The resulting Stochastic Search Variable Selection Expectation–Maximisation (SSVS–EM) algorithm provides the MaP of the inclusion probability of each regressor avoiding the computation of the whole posterior distribution. In Section 3, we introduce an Expectation–Maximisation algorithm for efficiently dealing with Maximum–a–Posteriori estimation of the relevant regression parameters as well as the parameters driving the variance of the error term. The postulated Spike-and-Slab prior on the endogenous regressors parameters β_i featuring the

presence of irrelevant regressors, i.e., $\beta_j = 0$ for some j in a probabilistic way, while retaining the common factor structure \mathbf{z}_t .

2.2 Network specification

For the sake of clarity, we rewrite the model in equation (4) to highlight its components in terms of network linkages. Indeed, the regression framework can be further simplified by considering that the regressors in the right-hand side contains the set of lagged variables denoted by $\mathbf{x}_{i,t}$ and the observed factors $\mathbf{z}_t = (y_{i,t-1}, y_{i,t-2}, \dots, y_{i,t-k}, \mathbf{f}_t^\top)^\top$. This latter set of variables is always included in the i -th regression while the former are selected by including a binary random vector γ_i of dimension $k(N-1)$ with $\gamma_{i,j} = 1$ if the j -th covariate $x_{i,j}$ is included as an explanatory variable in the i -th regression model and 0 otherwise.

Therefore, using standard notation, we include the selection variable $\gamma_{i,j} \sim \text{Ber}(\omega_i)$ in the previous specification, leading to the following representation:

$$y_{i,t} = \mu_i + \sum_{\substack{j=1 \\ j \neq i}}^{k(N-1)} \gamma_{i,j} x_{i,j,t-1} \beta_{i,j} + \mathbf{z}_t^\top \boldsymbol{\varphi}_i + h_{i,t}^{1/2} \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \mathbf{N}(0, 1), \quad (6)$$

$$h_{i,t} = \gamma_{i,h} + \alpha_{i,h} \varepsilon_{i,t-1}^2 + \beta_h h_{i,t-1}, \quad (7)$$

for $i = 1, 2, \dots, N$. The regression term $\boldsymbol{\varphi}_i = (\varrho_{i,1}, \varrho_{i,2}, \dots, \varrho_{i,k}, \mathbf{f}_t^\top)^\top$ refers to the common factor components which are always included by default in the model. The heteroscedastic regression framework specified in equations (6)–(7) can be used to infer a probabilistic network over the set of nodes $\mathcal{V} = \{1, 2, \dots, N\}$ with directed edges between nodes represented through a N -dimensional adjacency matrix

$$A = \begin{pmatrix} 0 & a_{1,2} & \cdots & a_{1,j-1} & a_{1,j} & \cdots & a_{1,N} \\ a_{2,1} & 0 & \cdots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \ddots & \cdots & \cdots & \cdots & \vdots \\ \vdots & \cdots & \cdots & \ddots & \cdots & \cdots & \vdots \\ a_{i,1} & \cdots & \cdots & a_{i,j-1} & a_{i,j} & \cdots & a_{i,N} \\ \vdots & \cdots & \cdots & \cdots & \cdots & \ddots & \vdots \\ a_{N,1} & \cdots & \cdots & a_{N,j-1} & a_{N,j} & \cdots & 0 \end{pmatrix}. \quad (8)$$

The element $a_{i,j} = g(\hat{\pi}_{i,j})$ is a function of the estimated inclusion probabilities

$$\hat{\pi}_{i,j} = \hat{\mathbf{P}}(\gamma_{i,j} = 1 \mid \mathbf{y}_i, \mathbf{X}_i, \mathbf{Z}_i, \mu_i, \boldsymbol{\beta}_i^\top, \boldsymbol{\varphi}_i^\top, \psi^\top), \quad (9)$$

where $\mathbf{y}_i = (y_{i,1}, y_{i,2}, \dots, y_{i,T})^\top$, $\mathbf{X}_i = (\mathbf{x}_{i,1}, \mathbf{x}_{i,2}, \dots, \mathbf{x}_{i,T})^\top$ and $\mathbf{Z}_i = (\mathbf{z}_{i,1}, \mathbf{z}_{i,2}, \dots, \mathbf{z}_{i,T})^\top$ denote the set of observations used to estimate the parameters $(\mu_i, \boldsymbol{\beta}_i^\top, \boldsymbol{\varphi}_i^\top, \psi^\top)^\top$.

The adjacency matrix A is symmetric and describes a weighted directed network since the linkages are defined in terms of inclusion probabilities and thus, the element $a_{i,j}$ indicates a (weighted) edge from j to i with $i, j \in \mathcal{V} = \{1, 2, \dots, N\}$. An alternative characterisation of the network is possible with the use of a threshold $c \in (0, 1)$ where the element $a_{i,j}$ of the adjacency matrix is defined as $g(\hat{\pi}_{i,j}) = \mathbb{1}_{(c,1)}(\hat{\pi}_{i,j})$, with $\mathbb{1}_{(\cdot)}(x)$ denoting the indicator function. In such case, there is a binary relationship: if $\hat{\pi}_{i,j} > c$, $a_{i,j}$ is equal to 1 and 0 otherwise. This can be further extend to multiple thresholds in order to have different linkage categories which depend on the probability weight of each edge (i.e., low–medium–high connectedness).

3 Heteroscedastic Spike–and–Slab SSVS–EM

In this section, we detail the EM algorithm to estimate the regression parameters of the model specified in equations (6)–(7), as well as their inclusion probabilities. For notational convenience, throughout this section we will suppress the dependence on the institution indicator i and assume $\boldsymbol{\beta} \in \mathbb{R}^p$ and $\boldsymbol{\varphi} \in \mathbb{R}^q$, with $p = k(N - 1)$ and $q = k + f$. The prior distribution for the inclusion parameters γ_j , $j = 1, 2, \dots, p$ as well as the other parameters $\mu, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \gamma_h, \alpha_h, \beta_h$ are specified in the next section. Section 3.2 introduces the EM algorithm that efficiently deals with the SSVS problem, while Section A.3 deals with the relevant problem of updating the GARCH parameters.

3.1 Prior specification

Using standard notation, let $\boldsymbol{\gamma}$ be the p –vector where $\gamma_j = 1$ if the j –th covariate x_j is included as an explanatory variable in the regression model and $\gamma_j = 0$ otherwise. Conditional on $\boldsymbol{\gamma}_j \sim \text{Ber}(\omega)$, the adaptive prior distribution for the remaining parameters are as follows:

$$\mu \sim N_1(\mu \mid \mu_\mu, \sigma_\mu^2) \quad (10)$$

$$(\boldsymbol{\beta} \mid \boldsymbol{\gamma}) \sim N_p(\boldsymbol{\beta} \mid 0, \boldsymbol{\Sigma}_\beta) \quad (11)$$

$$\boldsymbol{\varphi} \sim N_q(\boldsymbol{\varphi} \mid \boldsymbol{\mu}_\varphi, \boldsymbol{\Sigma}_\varphi) \quad (12)$$

$$\boldsymbol{\Sigma}_\beta = \text{diag}\{b_1, b_2, \dots, b_p\} \quad (13)$$

$$b_j = (1 - \gamma_j) \nu_0 + \gamma_j \nu_{1,j}, \quad 0 \leq \nu_0 < \min\{\nu_1\}, \quad j = 1, 2, \dots, p \quad (14)$$

$$\nu_1 \sim \prod_{j=1}^p \frac{\nu_{1,j}^b (1 - \nu_{1,j})^{-a-b-2}}{\text{Be}(a+1, b+1)} \mathbb{1}_{(0,\infty)}(\nu_{1,j}) \quad (15)$$

$$\omega \sim \frac{\omega^{a-1} (1 - \omega)^{b-1}}{\text{Be}(a, b)} \quad (16)$$

$$\gamma_h \sim \mathbf{N}_1(\gamma_h \mid \mu_\gamma, \sigma_\gamma^2) \mathbb{1}_{(0, \infty)}(\gamma_h) \quad (17)$$

$$\alpha_h, \beta_h \sim \mathbf{N}_1(\alpha_h \mid \mu_\alpha, \sigma_\alpha^2) \times \mathbf{N}_1(\beta_h \mid \mu_\beta, \sigma_\beta^2) \times \mathbb{1}_{\mathcal{S}}(\alpha_h, \beta_h), \quad (18)$$

where $b = 0$, $a \in (-1, 0]$ (see, e.g., [Cui and George, 2008](#), [Liang et al., 2008](#)) and \mathcal{S} denotes the simplex $\alpha_h \geq 0$, $\beta_h \geq 0$ and $\alpha_h + \beta_h < 1$ to ensure weekly stationarity of the process (see, e.g., [Francq and Zakoïan, 2010](#)). In equations (10)–(18), $\mathbf{N}_d(\cdot)$ denotes the Gaussian distributions of dimension $d \in \mathbb{Z}^*$ where \mathbb{Z}^* denotes the set of positive integers, while $\text{Ber}(\cdot)$ and $\text{Be}(\cdot)$ denote the Bernoulli and Beta distributions, respectively. Hereafter, $\boldsymbol{\vartheta} = (\mu, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top, \omega, \boldsymbol{\nu}_1, \psi^\top)^\top$ collects all the unknown parameters that need to be estimated.

The prior structure in equations (10) and (12) implicitly assumes that the constant term and the covariates \mathbf{z}_t , which correspond to the factors common to the whole system of VAR equations defined in the previous section, are always included as covariates. Therefore, we assume a Gaussian and multivariate Gaussian distribution, respectively. Regarding the prior distribution for the GARCH parameters $\psi = (\gamma_h, \alpha_h, \beta_h)^\top$ a natural solution is the independent Gaussian truncated over the simplex \mathcal{S} defined in equations (17)–(18) (see, e.g., [Ardia, 2008](#)). Equations (11) and (13)–(16) specify a hierarchical Spike-and-Slab prior for the vector of regressors $\boldsymbol{\beta}$. The Spike-and-Slab prior ([Mitchell and Beauchamp, 1988](#), [George and McCulloch, 1993, 1997](#)) is a finite mixture of two zero-mean Gaussian distributions and different variance, where the first component, with smaller variance, refers to the spike component while the second component, with greater variance, refers to the slab component. Therefore, the covariates that are excluded from the regression are assigned to the spike distribution, while the slab component accounts for those covariates that are included in the regression. The auxiliary $\boldsymbol{\gamma}$ vector of mixture indicator accounts for the inclusion of regressors enabling the data-augmentation approach for the Spike-and-Slab mixture. Conditionally on $\boldsymbol{\gamma}$, the joint prior distribution of $\boldsymbol{\beta}$ is Gaussian with mean zero and diagonal variance-covariance matrix $\boldsymbol{\Sigma}_\beta$ as specified in equation (12). The structure of the variance-covariance matrix in equations (13)–(14) allows the identification of the mixture components via the vector of prior variances $(\nu_0, \boldsymbol{\nu}_1^\top)^\top$. Even though the ν_0 in the spike distribution is typically set to be zero in the literature, such as in [Brown et al. \(1998\)](#) and [Panagiotelis and Smith \(2008\)](#), we follow [Ročková and George \(2014\)](#) and consider small but positive values for ν_0 to exclude unimportant nonzero effects. As suggested by [Ročková and George \(2014\)](#), we impose a heavy-tailed prior for $\boldsymbol{\nu}_1$ in equation (15) which is also known as a Pearson Type VI or Beta-prime distribution. The prior distribution specified in equations (14)–(15) is adaptive in the sense that it introduces a parameter specifically tailored to each regressor. Moreover, the hierarchical formulation allows posterior data-driven inference on $\boldsymbol{\nu}_1$.

The prior specification is completed by assuming a distribution for the hyper-parameter ω in equation (16).

The joint prior distribution specified in equations (10)–(18) depends on the following set of hyper-parameters $(\mu_\mu, \sigma_\mu^2, \boldsymbol{\mu}_\varphi, \boldsymbol{\Sigma}_\varphi, \nu_0, a, b, \mu_\gamma, \sigma_\gamma^2, \mu_\alpha, \sigma_\alpha^2, \mu_\beta, \sigma_\beta^2)$ that should be chosen to perform model estimation. As concerns the prior for the constant μ and the regressors that are always included $\boldsymbol{\varphi}$ we choose diffuse parameters, i.e., $\mu_\mu = 0$, $\sigma_\mu^2 = 100$ and $\boldsymbol{\mu}_\varphi = 0 \times \boldsymbol{\iota}_q$ and $\boldsymbol{\Sigma}_\varphi = 100 \times \mathbb{I}_q$ where $\boldsymbol{\iota}_q$ denotes a column vector of unit element of dimension q and \mathbb{I}_q denotes the identity matrix of order q . As discussed above, $b = 0$ while a common choice for the parameter a is $a = -\frac{3}{4}$, see [Maruyama and George \(2011\)](#). The hyper-parameter ν_0 is crucial because it is a tuning parameter that regulates the number of covariates that are set to zero. It can be either fixed to a small positive value (e.g., $\nu_0 = 10^{-6}$) or it can be selected by cross-validation. Although more computationally intensive we opt for the latter empirical Bayes approach, (see, e.g., [Park and Casella, 2008](#), [Cui and George, 2008](#), for an extensive treatment of an empirical approach to tailor hyper-parameters in Bayesian sparse regression models). The last group of hyper-parameters refers to the GARCH parameters. Following [Ardia \(2008\)](#), we set the means and the variances to $\mu_\gamma = \mu_\alpha = \mu_\beta = 0$ and $\sigma_\gamma^2 = \sigma_\alpha^2 = \sigma_\beta^2 = 100$, respectively.

3.2 The SSVS-EM algorithm

Hereafter, we focus on the EM algorithm which has been extensively applied in the context of finite mixture models, (see, e.g., [McLachlan and Peel, 2000](#)). For the covariates selection, the mixture representation originates from the Spike-and-Slab prior distribution placed on the regressor parameters in equations (10)–(18). To account for the missing information on the covariates included in the regression model, the vector of observations $\{y_t, \mathbf{x}_t, \mathbf{z}_t\}$ is regarded as being incomplete and the vector of indicators γ_t , that relies on the Spike-and-Slab mixture representation, is introduced as latent parameter.

This leads to the following complete-data log-likelihood:

$$\begin{aligned} \log \mathcal{L}_c(\boldsymbol{\vartheta}) = & -\frac{T+p+q+3}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log(h_t) - \frac{1}{2} \sum_{t=1}^T \frac{\|y_t - \mu - \mathbf{x}_t^\top \boldsymbol{\beta} - \mathbf{z}_t^\top \boldsymbol{\varphi}\|_2^2}{h_t} \\ & - \frac{1}{2} \sum_{j=1}^p \log(b_j) - \frac{1}{2} \sum_{j=1}^p \frac{\beta_j^2}{b_j} + |\gamma| \log\left(\frac{\omega}{1-\omega}\right) + p \log(1-\omega) \\ & - (a+b+2) \sum_{j=1}^p \log(1+\nu_{1,j}) + b \sum_{j=1}^p \log(\nu_{1,j}) - \sum_{j=1}^p \log(\text{Be}(a+1, b+1)) \end{aligned}$$

$$\begin{aligned}
& + (a - 1) \log(\omega) + (b - 1) \log(1 - \omega) - \log(\text{Be}(a, b)) - \frac{1}{2} \log(\sigma_\gamma^2) \\
& - \frac{1}{2} \frac{(\gamma_h - \mu_\gamma)^2}{\sigma_\gamma^2} - \frac{1}{2} \log(\sigma_\alpha^2) - \frac{1}{2} \frac{(\alpha_h - \mu_\alpha)^2}{\sigma_\alpha^2} - \frac{1}{2} \log(\sigma_\beta^2) - \frac{1}{2} \frac{(\beta_h - \mu_\beta)^2}{\sigma_\beta^2} \\
& - \frac{f + 3}{2} \log(2\pi) - \frac{1}{2} \log|\Sigma_\varphi| - \frac{1}{2} (\varphi - \mu_\varphi)^\top \Sigma_\varphi^{-1} (\varphi - \mu_\varphi), \tag{19}
\end{aligned}$$

where $\|\cdot\|_2^2$ denotes the squared L²-norm.

The EM algorithm consists of two major steps, one for expectation (E-step) and one for maximisation (M-step), see [McLachlan and Krishnan \(2007\)](#). At the $(m + 1)$ -th iteration the EM algorithm proceeds as follows:

- (i) **E-step:** computes the conditional expectation of the complete-data log-likelihood (19) given the observed data $\{y_t, \mathbf{x}_t, \mathbf{z}_t\}_{t=1}^T$ and the m -th iteration parameters updates $\hat{\boldsymbol{\vartheta}}^{(m)}$

$$\mathcal{Q}(\boldsymbol{\vartheta}, \hat{\boldsymbol{\vartheta}}^{(m)}) = \mathbb{E}_{\hat{\boldsymbol{\vartheta}}^{(m)}} \left[\log \mathcal{L}_c(\boldsymbol{\vartheta}) \mid \{y_t, \mathbf{x}_t, \mathbf{z}_t\}_{t=1}^T \right], \tag{20}$$

where the expectation in equation (20) is taken with respect the conditional distribution of the latent inclusion indicator γ given the observed variables $\{y_t, \mathbf{x}_t, \mathbf{z}_t\}_{t=1}^T$ and the current value of the parameters $\hat{\boldsymbol{\vartheta}}^{(m)}$;

- (ii) **M-step:** updates the parameters $\boldsymbol{\vartheta}$ by maximising equation (20) with respect to $\boldsymbol{\vartheta}$, as follows:

$$\hat{\boldsymbol{\vartheta}}^{(m+1)} = \arg \max_{\boldsymbol{\vartheta}} \mathcal{Q}(\boldsymbol{\vartheta}, \hat{\boldsymbol{\vartheta}}^{(m)}).$$

Given the GARCH parameters, the nonlinear regression model does not admit a closed form update for the M-step, and thus we resort to a Conditional Expectation Maximisation approach (CEM) (see [McLachlan and Krishnan, 2007](#)). The Heteroscedastic SSVS-EM algorithm included in Algorithm 1 describes the main steps for the linear regression model with Spike-and-Slab prior. The full derivation of the explicit EM algorithm is reported in the Appendix A.1.

4 Empirical Analysis

In this Section, we apply the proposed model to the context of the global financial system. The analysed period goes from 29 December 2000 to 6 October 2017. The data have been downloaded from Thomson Reuters Eikon. As commonly used in literature ([Billio et al., 2012](#), [Diebold and Yilmaz, 2014](#), [Demirer et al., 2017](#)), we make use of the rolling window approach to obtain a dynamic network over time. Therefore, the networks are estimated using a window with length

Algorithm 1 Heteroscedastic Stochastic Search Variable Selection EM

Given the value of the parameters at the m -th iteration, $\hat{\boldsymbol{\vartheta}}^{(m)}$ as well as the values of the conditional variances $\left\{\hat{h}_t^{(m)}\right\}_{t=1}^T$, the next iteration of the SSVS-EM algorithm update the parameters as follows:

1. update the inclusion indicators: $\hat{\gamma}_j^{(m)} = \frac{1}{1+\hat{d}_j^{(m)}}$, where

$$\hat{d}_j^{(m)} = \sqrt{\frac{\hat{\nu}_{1,j}^{(m)}}{\nu_0}} \exp \left\{ \frac{\left(\hat{\beta}_j^{(m)}\right)^2}{2} \left(\frac{1}{\hat{\nu}_{1,j}^{(m)}} - \frac{1}{\nu_0} \right) \right\} \frac{1 - \hat{\omega}^{(m)}}{\hat{\omega}^{(m)}},$$

for $j = 1, 2, \dots, p$;

2. update the latent factors:

$$\begin{aligned} \widehat{b_j^{-1}}^{(m)} &= \frac{1}{\nu_{1,j}} \hat{\gamma}_j^{(m)} + \frac{1}{\nu_0} (1 - \hat{\gamma}_j^{(m)}) \\ \widehat{\log(b_j)}^{(m)} &= \log(\nu_{1,j}) \hat{\gamma}_j^{(m)} + \log(\nu_0) (1 - \hat{\gamma}_j^{(m)}), \end{aligned}$$

3. update the regression parameters $\boldsymbol{\vartheta}_* = (\boldsymbol{\mu}, \boldsymbol{\beta}^\top, \boldsymbol{\varphi}^\top)^\top$:

$$\begin{aligned} \hat{\boldsymbol{\vartheta}}_*^{(m+1)} &= \boldsymbol{\mu}_{\vartheta_*} + \widehat{\mathbf{K}}_*^{(m)} (\mathbf{y} - \mathbf{X}_* \boldsymbol{\mu}_{\vartheta_*}) \\ \hat{\omega}^{(m+1)} &= \frac{1}{p + b + a - 2} \left(\sum_{j=1}^p \hat{\gamma}_j^{(m)} + a - 1 \right) \\ \hat{\nu}_1^{(m+1)} &= - \frac{\left(\widehat{A}^{(m)} + \widehat{B}^{(m)} + b \right) + \widehat{\Delta}^{(m)}}{2 \left(\widehat{B}^{(m)} - a - 2 \right)}, \end{aligned}$$

where $\widehat{\mathbf{K}}_*^{(m)}$, $\widehat{\Delta}^{(m)}$, $\widehat{B}^{(m)}$, $\widehat{A}^{(m)}$ are defined in equations (25), (29), (30), (31) provided in Appendix A, and $\boldsymbol{\mu}_{\vartheta_*} = (\boldsymbol{\mu}_\mu, \mathbf{0}^\top, \boldsymbol{\mu}_\varphi^\top)^\top$;

4. update the GARCH parameters $\hat{\psi}_h^{(m+1)} = \left(\hat{\gamma}_h^{(m+1)}, \hat{\alpha}_h^{(m+1)} \right)^\top$ and $\hat{\beta}_h^{(m+1)}$:

$$\begin{aligned} \hat{\psi}_h^{(m+1)} &= \left(\mathbf{L}^\top \boldsymbol{\Lambda}^{-1} \mathbf{L} + \boldsymbol{\Sigma}_{\tilde{\psi}_h}^{-1} \right)^{-1} \left(\mathbf{L}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\varepsilon} + \boldsymbol{\Sigma}_{\tilde{\psi}_h}^{-1} \boldsymbol{\mu}_{\tilde{\psi}_h} \right) \\ \hat{\beta}_h^{(m+1)} &= \left(\boldsymbol{\nabla}_\beta^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\nabla}_\beta + \frac{1}{\sigma_{\beta_h}^2} \right)^{-1} \left(\boldsymbol{\nabla}_\beta^\top \boldsymbol{\Lambda}^{-1} \mathbf{r} + \frac{\boldsymbol{\mu}_{\beta_h}}{\sigma_{\beta_h}^2} \right), \end{aligned}$$

where \mathbf{L} , $\boldsymbol{\Lambda}$, \mathbf{r} and $\boldsymbol{\nabla}_\beta$ are defined in equations (66)–(69) of Appendix A.

of 104 weekly observations (2 stock market years) for a total of 772 adjacency matrices.¹ For each window the adjacency matrix $A \equiv A(t)$ is estimated using inclusion probabilities $\hat{\pi}_{i,j} \equiv \hat{\pi}_{i,j}(t)$ defined in equation (8) and (9). The prior parameters are chosen as described in Section 3.1. Furthermore, to cross-validate the tuning parameter ν_0 , we implement a warm-start technique which uses the current solution of ν_0 as the initial value for the next one (Friedman et al., 2007,

¹The estimations have been parallelised and implemented in Matlab on the LOEWE-CSC cluster using 100 cores. The computation has required approximately 72 hours.

2010, Zhao et al., 2018).

4.1 The data

Financial firms. We consider the weekly closing price series (friday) of 1248 world financial firms (banks, brokers insurances and other financial services). We account for both active and dead financial firms in order to deal with delisting in stocks markets and avoid the survivorship bias (Shumway, 1997). We report the number of financial firms per country and category in Table 5 of Appendix B.² Each estimation window contains financial firms with at least a number of observations equal to its size. Figure 1 reports the cardinality of the dataset (n_t) over time with a minimum (maximum) of selected firms equal to 766 (1178). Therefore, we have a time-varying dataset where the number of covariates are strictly larger than the number of observations.

Common observed factors. We select 44 world financial variables which may exert an impact on the cross-section correlation as described in equation 2. These variables include:

- *The Fama and French five global factors (Fama and French, 2015):* *i)* Mkt-Rf (the market's excess return); *ii)* SMB (Small Minus Big); *iii)* HML (High Minus Low); *iv)* RMW (Robust Minus Weak) and *v)* CMA (Conservative Minus Aggressive)³.
- *Log-returns on MSCI market indices:* Global, United States, Europe, China, Japan, Hong Kong and Russia. (Source: Bloomberg).
- *Changes volatility indices:* VIX, EUROSTOXX 50 Volatility Index (V2X), Nikkei Volatility Index (VNKY) and the Hong Kong Volatility Index (VHSI). (Source: Bloomberg).
- *Log-returns on bilateral spot exchange rate.* The rates are expressed in units of foreign currency per USD and include: Australia, Euro, China, Japan, Hong Kong, Norway, Russia and Sweden. (Source: Bloomberg).
- *Log-returns on 10 Years Government bond indices:* Austria, Australia, Bulgaria, China, Denmark, Finland, France, Germany, Greece, Ireland, Italy, Japan, Netherlands, Norway, Portugal, Spain, United Kingdom and United States. (Source: Eikon/Datastream).

²The dataset is part of the earliness.eu project (H2020) and of the SAFE Systemic Risk Lab, Goethe University Frankfurt.

³The Fama and French five factors makes use of six value-weight portfolios formed on size and book-to-market, the six value-weight portfolios formed on size and operating profitability, and the six value-weight portfolios formed on size and investment. For further information, see http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5_factors_2x3.html.

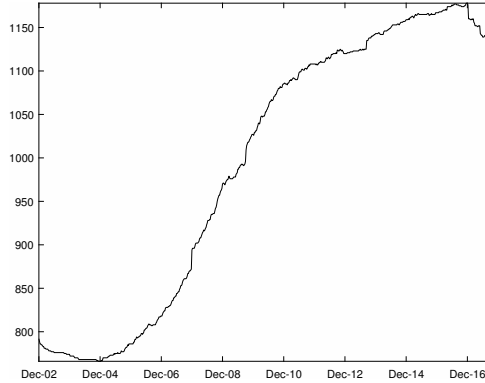


Figure 1: Cross-sectional sample size (n_t) of the considered financial institutions over time.

- *The TED spread* (3-month USD LIBOR minus the 3-month US Treasury bill yield) as a measure of counterparty risk and liquidity of the US interbank market and log-returns on *WTI Crude oil*.

4.2 Network estimates

In this application, we consider the binary adjacency matrix case and thus select the regressors with a probability of inclusion, $\hat{\pi}_{i,j}$, higher than a given threshold parameter, c . Rořková and George (2014) suggest considering threshold values larger than 0.5 to have more sparsified models. In this respect, we set c equal to 0.8 to proxy highly likely financial linkages among firms. Figure 2 reports the cross-sectional distribution over time for the number of selected regressors of each financial firm (left panel) and the cross-sectional distribution over time of the estimated betas for the financial firms (right panel). Looking at the left panel, the cross-sectional distribution of the selected regressors provides a first indication about the distribution of the ingoing connections among the financial firms in the system. Interestingly, there is a shift on the the cross-sectional distribution during the 2007–2009 period in occurrence of the Global financial crisis which could indicate a structural change in the network connectedness. In this regard, we report in Table 1 the descriptive statistics for the cross-sectional distributions of the selected regressors per year. The values for the mean are larger than the values of the median in all of the considered period which suggests that the cross-sectional distribution through time could be right-skewed. As shown graphically, there is a decline of the mean (first column) and the median (second column) during the global financial crisis, even if the latter is less pronounced than the former. For a given financial firm, a decrease of the selected regressors on average indicates a lower number of causal relationships, a signal that the shock transmission mechanism is primarily driven by fewer institutions. The maximum difference

between the mean and the median is during 2009–2010 and indicates that the distribution is strongly right-skewed in the period which suggests the presence of a minority of nodes with a large number of connections. Conversely, the standard deviation (third column) increases by 42% in 2007 indicating a greater dispersion among the number of ingoing connections. This can be also viewed on the quantiles at 10% (fourth column) and 90% (fifth column) which indicate a widening of the distribution during the same period (2009–2010). The right panel in Figure 2 shows a bimodal cross-sectional distribution for the betas with a mass concentration around positive and negative values equal to 0.15 and -0.15 . The descriptive statistics are reported in Table 2 and show that the mean, median and standard deviation for the betas are substantially stable during the whole period.

To analyse the shock transmission mechanism, we focus on the outgoing linkages in the network over time. The out-degree measure, $d_{i,t}^{\text{out}}$, is defined as

$$d_{i,t}^{\text{out}} = \sum_{j=1}^{n_t} a_{ij,t},$$

where $a_{ij,t}$ is the element $\mathbb{1}_{(c,1)}(\hat{\pi}_{i,j})$ of the adjacency matrix at time t . Figure 3 reports the cross-sectional out-degree distribution over time for the considered financial firms. Similarly as discussed for the selected regressors, the cross-sectional out-degree distribution over time is right-skewed indicating the presence of a majority of nodes having low out-degrees and a minority also known as financial hubs with high out-degrees. This is a well known property of scale-free networks where the degree distribution is usually heavy tailed (Barabási and Albert, 1999) and follows a power tail distribution, a characteristic also present in financial networks (see Iori et al., 2008, Schweitzer et al., 2009). From a financial stability perspective, financial hubs are key players on the shock transmission in the system given their central role played in the network (Boss et al., 2004, Haldane, 2013). As shown in Billio et al. (2012) and Diebold and Yilmaz (2014), changes in network connectedness, measured by the total density, the spillover index and other network indicators, can provide valuable signals about periods of stress and instability in the financial system.

In this respect, we consider the cross-sectional variation of the out-degrees over time and analyse the shape of the distribution as a potential indicator of market stress. We make use of the Generalized Pareto Distribution (GPD) (Pickands, 1975) which is particularly suitable for heavy-tailed distributions and is also used to model network degrees (Jordan et al., 2004, Barabási, 2016).

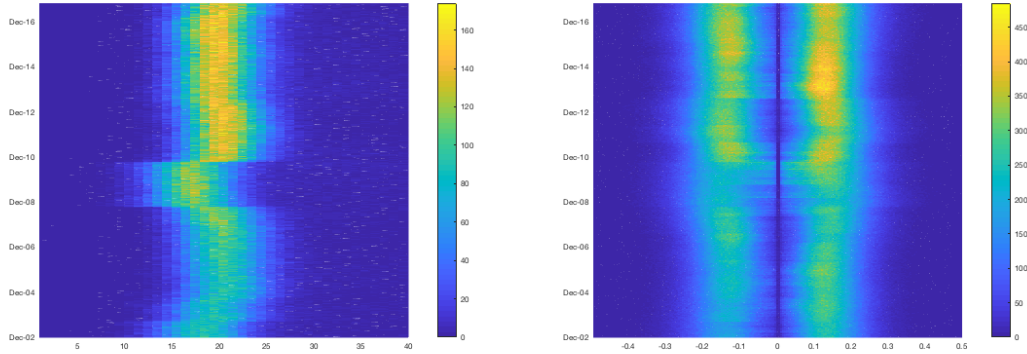


Figure 2: Cross-sectional distribution of the selected regressors for $c > 0.8$ (left panel) and cross-sectional distribution for estimated betas (right panel) from 20 December 2002 to 6 October 2017. The colour indicates the intensity of the frequency from blue to yellow (see the colourmap).

	<i>mean</i>	<i>median</i>	<i>std. dev.</i>	<i>q</i> _{10%}	<i>q</i> _{90%}
<i>2002</i>	18.01	17.00	8.23	12.50	21.50
<i>2003</i>	19.30	17.54	10.28	13.23	23.13
<i>2004</i>	21.51	18.82	12.35	14.57	25.79
<i>2005</i>	22.24	19.71	12.26	15.53	26.37
<i>2006</i>	21.08	19.16	10.22	15.11	24.78
<i>2007</i>	20.79	19.00	14.51	15.05	24.06
<i>2008</i>	20.20	17.96	11.00	14.02	24.03
<i>2009</i>	20.07	16.00	15.32	12.04	26.57
<i>2010</i>	21.86	17.00	17.03	12.85	32.93
<i>2011</i>	21.63	19.12	11.39	15.38	25.69
<i>2012</i>	20.50	19.21	8.40	15.32	23.96
<i>2013</i>	20.96	18.94	11.07	14.88	25.03
<i>2014</i>	21.23	18.96	12.06	14.94	25.26
<i>2015</i>	21.18	18.98	11.61	14.97	24.79
<i>2016</i>	20.60	18.47	11.11	14.64	24.02
<i>2017</i>	20.77	18.96	9.99	15.00	24.33

Table 1: Yearly descriptive statistics (mean, median, standard deviation and the 10th and 90th percentile) for the cross-sectional distribution of the number of the selected regressors for each institution.

	<i>mean</i>	<i>median</i>	<i>std. dev.</i>		<i>mean</i>	<i>median</i>	<i>std. dev.</i>
<i>2002</i>	-0.1521	-0.1475	0.0711	<i>2010</i>	-0.1391	-0.1337	0.0787
	0.1462	0.1400	0.0674		0.1408	0.1341	0.0762
<i>2003</i>	-0.1477	-0.1427	0.0727	<i>2011</i>	-0.1430	-0.1381	0.0717
	0.1431	0.1377	0.0682		0.1492	0.1445	0.0715
<i>2004</i>	-0.1404	-0.1357	0.0716	<i>2012</i>	-0.1539	-0.1486	0.0727
	0.1396	0.1347	0.0679		0.1571	0.1517	0.0718
<i>2005</i>	-0.1358	-0.1311	0.0693	<i>2013</i>	-0.1452	-0.1403	0.0726
	0.1371	0.1320	0.0666		0.1431	0.1380	0.0684
<i>2006</i>	-0.1406	-0.1355	0.0684	<i>2014</i>	-0.1367	-0.1328	0.0691
	0.1418	0.1365	0.0660		0.1363	0.1315	0.0657
<i>2007</i>	-0.1425	-0.1373	0.0679	<i>2015</i>	-0.1373	-0.1326	0.0692
	0.1521	0.1466	0.0690		0.1405	0.1359	0.0678
<i>2008</i>	-0.1512	-0.1469	0.0742	<i>2016</i>	-0.1432	-0.1382	0.0709
	0.1543	0.1497	0.0738		0.1472	0.1418	0.0710
<i>2009</i>	-0.1509	-0.1476	0.0808	<i>2017</i>	-0.1447	-0.1400	0.0709
	0.1513	0.1463	0.0788		0.1488	0.1431	0.0715

Table 2: Yearly cross-sectional descriptive statistics (mean, median and standard deviation) on the betas of the selected regressors.

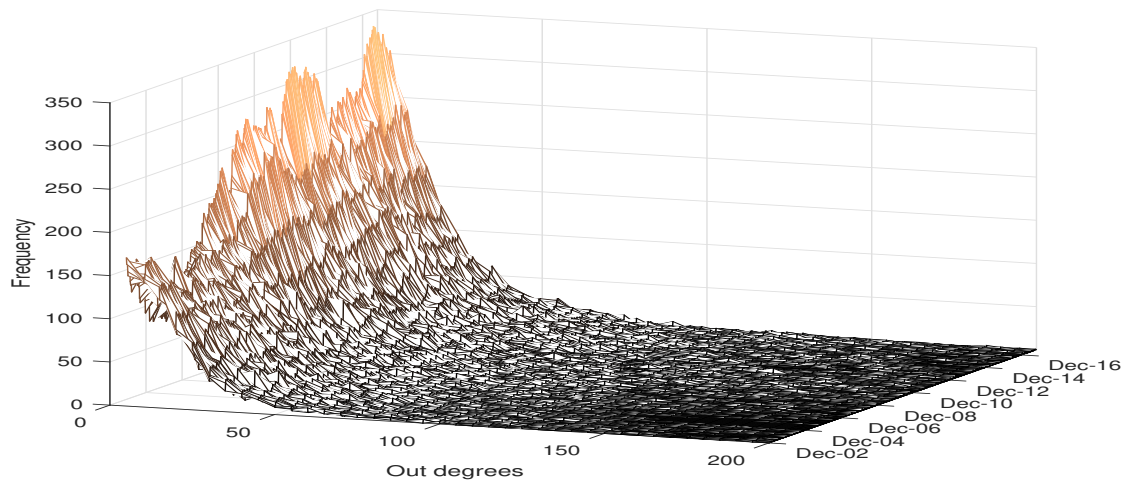


Figure 3: Cross-sectional out-degree distribution of the considered financial institutions over time. The frequency (y-axes) indicates the number of financial firms with the correspondent number of out-degrees (x-axes).

The cumulative distribution function of the GPD is

$$F_{x,\xi,\beta} = \begin{cases} 1 - \left(1 + \xi \frac{x}{\beta}\right)^{-1/\xi}, & \text{if } \xi \neq 0, \beta > 0, \\ 1 - e^{-x/\beta}, & \text{if } \xi = 0, \beta > 0. \end{cases}$$

where β is the scale parameter and ξ represents the shape parameter (tail index) which models the weight of the tail.⁴

We provide in Figure 4 a simple intuition as to why the variation over time of the shape parameter ξ can represent a meaningful signal of structural changes on the network connectedness. *Ceteris paribus*, an increase of the shaper parameter ξ at time t would result in a more heavy-tailed degree distribution (Panel b) with respect to the baseline scenario in $t - 1$ (Panel a). Such a change would indicate an increase in the number of nodes with higher out-degrees (red bars in the figure) due to either the formation of new hubs or a strengthening of the existing ones.

Indeed, we estimate the GPD on the out-degree distribution over time and obtain a time-sequence of the shape parameter $\hat{\xi}_t$. For the sake of simplicity, we name this sequence hereafter the Hubs Indicator (HIND) since an increase over time in the shape of the degree distribution represents a signal of a higher number of financial hubs, that is, nodes with higher (out-going) connections in the financial system. The HIND is included in Figure 5 and shows an interesting dynamic with occasional level shifts in occurrence of particular financial and economic events before and after the global financial crisis and the subsequent European debt crisis.⁵ Given that the two crises are intimately associated, we identify the main jumps on the HIND and report the correspondent episodes in red dash-dotted lines plus two more recent episodes that influenced the global financial market listed in Table 3. It is worth noting that there is an increase in the HIND during the first part of the sample as indicated by the solid black line which goes from the global minimum in June 2005 to the global maximum in September 2010. Before the timeline events of the global financial crisis there are two type of episodes that are detected by the indicator. The first type is a dramatic weather event: the Hurricane Katrina which occurred on 29 August 2005 (#1) and had a significant impact on the energy market, especially on crude oil and natural gas prices (Brown and Yücel, 2008, Geman, 2009). As a direct consequence of this, there were severe financial losses for investment funds such as energy hedge funds (Fusaro, 2005). The second type involves monetary policy decisions by the FED

⁴Note that, if $\xi = 0$, the GPD distribution reduces to the exponential distribution with mean β . Moreover, $x \geq 0$, if $\xi \geq 0$ and $0 \leq x \leq -\beta/\xi$ otherwise. For further details, see for example Castillo and Hadi (1997) and Embrechts et al. (2013).

⁵We have checked that the condition $0 \leq x \leq -\beta/\xi$ is met for negative values of $\hat{\xi}_t$. Moreover, we test that the dynamic of the shape indicator $\hat{\xi}_t$ is not affected by the time-varying cross-sectional sample size reported in Figure 1 by performing a regression on levels and differences.

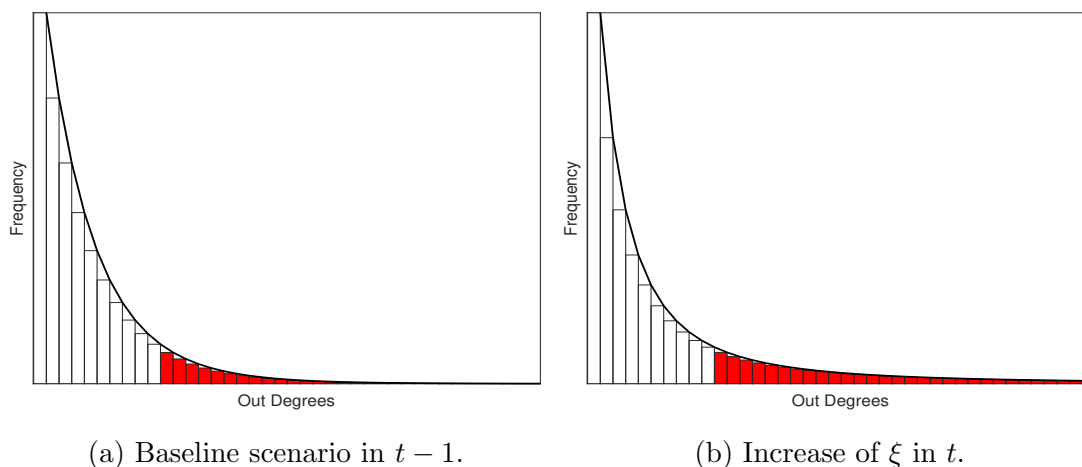


Figure 4: Example of an increase of the shape parameter ξ in t (Panel b) with respect to the baseline scenario in $t - 1$ (Panel a). Such a structural change in the network connectedness indicates an increase in the number of nodes with higher out-degrees (red bars) due to either the formation of new hubs or a strengthening of the existing ones.

#	Date	Event	Description
1	Aug 2005	Hurricane Katrina	Impact on the commodities market.
2	May 2006	FED 16th consecutive raise of interest rates	Inflation pressure with concerns for the global economic growth.
3	Jun 2006	FED 17th consecutive raise of interest rates	
4	Feb 2007	HSBC announces losses due to the US subprime mortgage market	Signal of vulnerability on the subprime mortgage market.
5	July 2007	Bear Stearns' two subprime Hedge Funds lost almost all the cap.	
6	Sep 2008	Bankruptcy of Lehman Brothers	Subprime financial crisis.
7	Jul 2011	Greece second bailout	European Sovereign debt crisis.
8	Jul 2013	30-Year US Treasury Swaps Spread returns positive	Measure of the cost of funding. Signal of a market normalization.
9	Dec 2015	Crude oil prices fall	WTI futures fell at \$36.66 per barrel.
10	Jul 2017	Global growth	Signals of continued and synchronized global growth.

Table 3: Dating of significant financial and economic events before and after the Global Financial crisis and the European Sovereign debt crisis plus two more recent episodes that influenced the global financial market. The episodes are reported in Figure 5 (dash-dot red lines).

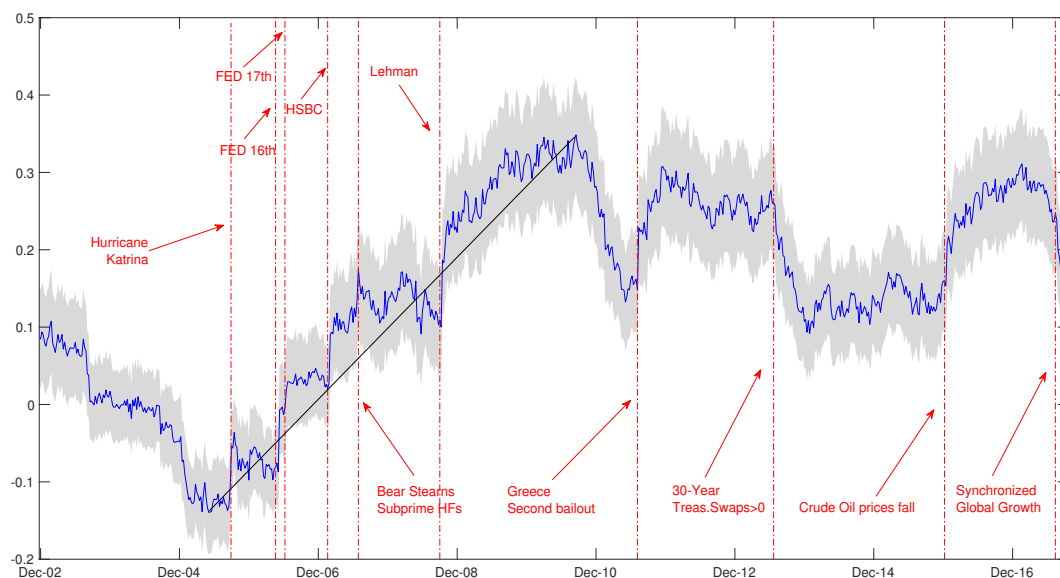


Figure 5: The Hubs Indicator (HIND) over time including the confidence interval at 95% (grey area). Red dash-dotted lines indicate financial and economic events as reported in Table 3.

with the 16th (#2) and 17th (#3) consecutive raise of interest rates in May and June 2006, respectively. Such decisions have been made to contrast inflation pressure and to cool housing market prices⁶ which in turn provoked concerns for global economic growth and generated a sell-off in particular on the emerging markets.⁷

With the beginning of 2007, the HIND shows a jump in correspondence with HSBC's announcement about financial losses on the subprime mortgage market (#4). At that time, HSBC was the third-largest bank in the world⁸ and one of the main players in the US subprime mortgage market. Then, the indicator continues to growth upwards with a second jump on the collapse of the two Bear Stearns' subprime hedge funds (#5). Both of these episodes represented a strong signal of the increasing vulnerability of the subprime mortgage market. The highest upward shift of the HIND indicator is on the Lehman bankruptcy (#6) in mid September 2008 (+128%), the tipping point for the global financial market crash which in turn provoked

⁶Statements by the Federal Open Market Committee (FOMC) on the 16th and 17th raise of interest rates are available at <https://www.federalreserve.gov/newsevents/pressreleases/monetary20060510a.htm> and <https://www.federalreserve.gov/newsevents/pressreleases/monetary20060629a.htm>.

⁷See for example the IMF release about the financial market situation available at <https://www.imf.org/External/Pubs/FT/fmu/eng/2006/0606.pdf>.

⁸At the time of writing, HSBC is the seventh world largest bank in terms of market capitalisation (Source: S&P Global Market Intelligence).

a generalised panic and global sell-off escalating into a liquidity crisis. The indicator becomes stable during the second part of 2009 and then starts to decrease to the pre-Lehman level at the end of May 2011. After that, it is clearly observable that there is a new upward shift in the indicator during the worsening of the European sovereign debt crisis with the second bailout of Greece (#7) in July 2011. The HIND peaks again on December 2011 corresponding to the most acute phase of the sovereign debt crisis of Italy and Spain, and then begins to decay until it stabilises during the second part of 2012. At the beginning of the second half of 2013, the indicator decreases given the signals of market normalization (#8) and reaches the pre-Lehman level at the beginning of 2014. Finally, the last upward shift of the HIND is in December 2015 during the fall of the crude oil prices (#9) which is also related to the slowdown of the Chinese economy. The indicator begins to decrease in 2017 with a downward shift in July when all the major financial markets had a strong positive performance due to the signals of synchronised global growth (MSCI World Index, +3.92%).

Interestingly, the dynamic of the HIND indicator is consistent with the cyclical variation of the system-wide connectedness presented in [Demirer et al. \(2017\)](#) which detected significant events during the Global financial crisis and the European sovereign debt crisis. In their analysis, the authors cover a period from 2004 to 2013 and found that the trend on the system-wide connectedness reaches its peak and starts to decline after the default of Lehman. Even if, in our considered case, the largest single upward shift occurs after the Lehman failure, the HIND reaches its peak on September 2010 and decreases immediately after the FED's announcement of a new round of quantitative easing (QE2) in November 2010. Both the indicators show an increase of interconnectedness and a peak during the worsening of the European sovereign debt crisis in mid 2011 with a decay immediately after. The HIND indicator also shows a new bump by mid-2015 which clearly cannot be compared with the system-wide connectedness of [Demirer et al. \(2017\)](#) for the sample length limitation in their analysis. However, this is visible on the total-connectedness index in [Korobilis and Yilmaz \(2018\)](#) obtained using the same variance decomposition approach through a time-varying parameter Bayesian VAR for 35 US and European financial institutions from January 2004 to July 2016.

Similarly with the financial stress indicators such as the St. Louis Fed Financial Stress Index (STLFSI, [Kliesen et al., 2010](#)) and the Composite Indicator of Systemic Stress (CISS, [Hollo et al., 2012](#)), the HIND proves to be very reactive to stress events. It is worth noting that, unlike those stress indices, measures such as the HIND, the spillover index and other network measures exploit the dependence structure among financial firms and are not obtained through the aggregation of different market sources (i.e., equity, bond, derivatives and forex markets).

As a further analysis, we test if the time variation of the HIND can represent a meaningful market predictor. We select as the proxy for the considered global market the MSCI World

Financials Index which captures large and mid cap representation across 23 developed countries and therefore, is fully consistent with the considered dataset.⁹ We model the market returns as a function of the usual market drivers (i.e., volatility and risk premia) including changes on the HIND (ΔHIND). Hence, we consider the Exponential-GARCH in mean model (EGARCH-m) which takes into account time varying risk premia in the mean equation and the asymmetry in the volatility equation:

$$r_t = c + \sum_{l=1}^L \phi_l r_{t-l} + \sum_{m=0}^M \delta_m h_{t-m}^{1/2} + \sum_{k=0}^K \beta_k \Delta\text{HIND}_{t-k} + h_t^{1/2} z_t,$$

$$\log(h_t) = \omega + \sum_{i=1}^q g(z_{t-i}) + \sum_{j=1}^p \phi_j \log(h_{t-i}),$$

where r_t denotes the MSCI index log-return at time t and $g(z_t) = \alpha z_t + \kappa [|z_t| - E|z_t|]$ and ΔHIND_{t-k} represents the changes in the indicator included in the mean equation at lag k . As usually considered in the literature, we limit the model orders to one, AR(1)-EGARCH-m(1,1), and include HIND_{t-k} up to the sixth lag ($k = 6$) to check its effect on market returns. The estimates for the mean and volatility equations are reported in Table 4. At a 5% confidence level, the autoregressive term r_{t-1} and the lagged (log) volatility $\log(h_{t-i})$ in the mean equation do not have any impact on the dynamic of the market returns while, as expected, all the terms in the volatility equation are significant. The HIND shows an impact on the weekly market returns at the first and the fourth lag which can be interpreted as a weekly (one week) and monthly effect (four week). In particular, the coefficient of ΔHIND_{t-1} (-0.1475) indicates that an increase of the shape indicator has a negative impact on the market returns at the first lag, while at the fourth lag, HIND_{t-4} , it has a positive impact (0.1115). We believe that this weekly and monthly effect are a reflection of heterogeneous market participants which differently affect the short and longer time horizons as described in Müller et al. (1997) and Corsi (2009) on the asymmetric propagation of volatility. In this respect, the negative sign for HIND_{t-4} may reflect the asymmetric impact of the monthly participant component over the weekly one resulting in a mean reverting effect on the market returns. As a robustness check, we assess if the HIND maintains its explanatory power on the market returns at the first and fourth lag after including another set of predictors: i) changes in volatility indices for four macro-areas which are representative of the global market index; ii) changes in financial stress indicators such as

⁹The MSCI World Financial index contains 23 developed countries for three main areas: i) *Americas*: Canada, United States; ii) *Europe&Middle East*: Austria, Belgium, Denmark, Finland, Germany, Ireland, Israel, Italy, Netherlands, Norway, Portugal, Spain, Sweden, Switzerland, United Kingdom; iii) *Pacific*: Australia, Hong Kong, Japan, New Zealand and Singapore. These countries represent more than the 62% of our dataset.

	estimate	std. err.	t-stat	p-value
<i>mean equation</i>				
(constant)	0.0010	0.0010	1.0000	0.3170
r_{t-1}	-0.0352	0.0402	-0.8800	0.3810
$\log(h_{t-1})$	0.3697	1.5314	0.2400	0.8090
ΔHIND_{t-1}	-0.1475	0.0333	-4.4300	0.0000
ΔHIND_{t-2}	-0.0948	0.0505	-1.8800	0.0600
ΔHIND_{t-3}	-0.0282	0.0446	-0.6300	0.5270
ΔHIND_{t-4}	0.1115	0.0466	2.3900	0.0170
ΔHIND_{t-5}	0.0326	0.0445	0.7300	0.4630
ΔHIND_{t-6}	-0.0045	0.0430	-0.1000	0.9170
<i>volatility equation</i>				
(constant)	-0.1540	0.0482	-3.2000	0.0010
z_{t-1}	-0.1363	0.0138	-9.8700	0.0000
$ z_{t-1} - E z_{t-1} $	0.0667	0.0228	2.9300	0.0030
$\log(h_{t-1})$	0.9798	0.0064	152.0000	0.0000
obs	765		Wald $\chi^2(5)$	29.25
Log-likelihood	1779.78		Prob $> \chi^2$	0.0003

Table 4: Estimates for the AR(1)–EGARCH–m(1,1) model by including the (lagged) changes in the Hubs Indicator ΔHIND_t in the mean equation. The considered period is from 20 December 2002 to 6 October 2017.

the CISS and the STLFSI and iii) the Fama and French 5 factors (Fama and French, 2015).¹⁰ The results are included Appendix C and confirm the previous findings for the HIND indicator in all the considered regressions.

5 Conclusion

Estimation of high-dimensional dynamic models is a relevant statistical issue that finds large application in econometric and finance in several contexts. Among those, systemic risk assessment plays a major role and strongly benefits from the availability of models that account for the evolution over time of the contagion mechanism within a network where vertexes represent financial firms. However, the curse of dimensionality problem that affects dynamic models, is that it limits their applicability to small-to-medium contexts even in their simpler linear autoregressive form. In this paper we deal with the problem of estimating high-dimensional VAR models by following a more realistic equation-by-equation estimation approach that effectively reduces the number of parameters to be estimated for each cross-sectional dimension. Moreover, since the main focus is to provide risk assessment based on networks, we cast the problem within the Bayesian framework that displays its power by allowing joint parameters estimation and model inference. Indeed, the model selection issues are related to the inclusion of covariates that can be interpreted as connections among vertexes. More specifically, we propose a shrinkage and selection methodology designed for network inference in high-dimensional data and present a regularised linear model with Spike-and-Slab prior on the regression coefficients (e.g., Mitchell and Beauchamp, 1988, George and McCulloch, 1993, 1997) where the marginal inclusion probability of each regressor can be obtained as a byproduct of the estimation procedure. To avoid the computational burden of MCMC methods, we follow the idea of Ročková and George (2014) and developed a fast and efficient Expectation-Maximisation (EM) algorithm (Dempster et al., 1977). In particular, we extend the SVSS-EM approach to the case where the error terms of the Gaussian linear regression model are heteroscedastic by adding an ARCH-type equation to model the dynamic evolution of the variance through an approximate-EM algorithm.

The proposed model can be interpreted as a VAR model with a diagonal variance-covariance matrix where shocks are instantaneously uncorrelated and the number of covariates is strictly larger than the number of observations. More interestingly, to account for cross-sectional dependence, we include in the regression additional observed factors that cannot be excluded from the regression. Therefore, the model accounts for two sets of covariates: the first set contains the predetermined variables which cannot be penalised (that is, the set of common factors and

¹⁰Data source in parenthesis: CISS(ECB-sdw), STLFSI (FRED-St.Louis Fed), Volatility indices (Bloomberg) and Fama and French 5 factors (French's website).

the autoregressive component of the given financial institution) while the second set contains all the considered financial firms. For a given institution, the financial linkages are obtained by the posterior inclusion probability which consequently return a weighted directed network.

In the empirical application, we consider the weekly closing price series of 1248 world financial firms active and dead from 29 December 2000 to 6 October 2017. We estimate dynamic networks using a rolling window approach and show that the shape of the out-degree distribution over time, namely the hubs indicator (HIND), exhibits the typical behaviour of financial stress indicators showing occasional level-shift during particular economic and financial events. Moreover, the HIND proved to be a significant predictor of global market returns at the first lag (one week) and the fourth lag (one month).

We believe that these results highlight the value of our proposed methodology in the inference of high-dimensional networks and can represent a useful tool among the ones presented in the literature. As wisely stated in [Diebold and Yilmaz \(2014\)](#), all these proposed methodologies “have much to learn from each other” and each can contribute to successfully measuring financial risks. Our approach is similar to [Billio et al. \(2012\)](#) for the interpretation of causal linkages but the connections are originated through the inclusion probabilities which provide a weighted financial network that is not based on variance decompositions and does not need any identifying assumption, unlike [Diebold and Yilmaz \(2014\)](#). Several other types of network analyses can be carried out with the presented model, such as community detection and clustering and by also considering other markets (corporate bonds). We leave this aspect for further research.

Acknowledgements

Michele Costola acknowledges financial support from the Marie Skłodowska-Curie Actions, European Union, Seventh Framework Program HORIZON 2020 under REA grant agreement n.707070 and the research support from the Research Center SAFE, funded by the State of Hessen initiative for research LOEWE. This paper benefited from comments and discussions with Massimiliano Caporin, Roberto Casarin, Lorenzo Frattarolo, Otmar Issing, Jan-Pieter Krahnen, Roberto Panzica, Loriana Pelizzon and the participants at the EARLINESS.eu Workshop at SAFE (Goethe University Frankfurt), the Mathematical and Statistical Methods for Actuarial Sciences and Finance 2018 (MAF, Madrid) and the Computational and Financial Econometrics 2017 (CFE, London). The usual disclaimers apply.

References

- Acemoglu, D., Ozdaglar, A., and Tahbaz-Salehi, A. (2015). Systemic risk and stability in financial networks. *American Economic Review*, 105(2):564–608.
- Acharya, V., Engle, R., and Richardson, M. (2012). Capital shortfall: A new approach to ranking and regulating systemic risks. *American Economic Review*, 102(3):59–64.
- Ahelegbey, D. F., Billio, M., and Casarin, R. (2016). Bayesian graphical models for structural vector autoregressive processes. *Journal of Applied Econometrics*, 31(2):357–386.
- Allen, F. and Babus, A. (2009). Networks in Finance. *The network challenge: strategy, profit, and risk in an interlinked world*, 367.
- Ardia, D. (2008). *Financial risk management with Bayesian estimation of GARCH models*, volume 612 of *Lecture Notes in Economics and Mathematical Systems*. Springer-Verlag, Berlin.
- Awartani, B. M. and Corradi, V. (2005). Predicting the volatility of the s&p-500 stock index via garch models: the role of asymmetries. *International Journal of Forecasting*, 21(1):167–183.
- Barabási, A.-L. (2016). *Network Science*. Cambridge University Press.
- Barabási, A.-L. and Albert, R. (1999). Emergence of scaling in random networks. *Science*, 286(5439):509–512.
- Barabási, A.-L. and Bonabeau, E. (2003). Scale-free networks. *Scientific American*, 288(5):60–69.
- Barigozzi, M. and Brownlees, C. (2019). Nets: Network Estimation for Time Series. *Journal of Applied Econometrics*.
- Bianchi, D., Billio, M., Casarin, R., and Guidolin, M. (2019). Modeling contagion and systemic risk. *Journal of Econometrics (forthcoming)*.
- Bien, J. and Tibshirani, R. J. (2011). Sparse estimation of a covariance matrix. *Biometrika*, 98(4):807–820.
- Billio, M., Caporin, M., Frattarolo, L., and Pelizzon, L. (2018). Networks in risk spillovers: a multivariate GARCH perspective. *Working paper*.
- Billio, M., Caporin, M., Panzica, R. C., and Pelizzon, L. (2016). The impact of network connectivity on factor exposures, asset pricing and portfolio diversification. *Working paper*.

- Billio, M., Casarin, R., and Rossini, L. (2019). Bayesian Nonparametric sparse VAR models. *Journal of Econometrics* (forthcoming).
- Billio, M., Getmansky, M., Lo, A., and Pellizon, L. (2012). Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 101:535–559.
- Bluhm, M. and Krahnen, J. P. (2014). Systemic risk in an interconnected banking system with endogenous asset markets. *Journal of Financial Stability*, 13:75–94.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*, 31(3):307–327.
- Boss, M., Elsinger, H., Summer, M., and Thurner 4, S. (2004). Network topology of the interbank market. *Quantitative finance*, 4(6):677–684.
- Brown, P. J., Vannucci, M., and Fearn, T. (1998). Multivariate Bayesian variable selection and prediction. *J. R. Stat. Soc. Ser. B Stat. Methodol.*, 60(3):627–641.
- Brown, S. P. and Yücel, M. K. (2008). What drives natural gas prices? *The Energy Journal*, pages 45–60.
- Bühlmann, P. (2006). Boosting for high-dimensional linear models. *The Annals of Statistics*, 34(2):559–583.
- Castillo, E. and Hadi, A. S. (1997). Fitting the generalized Pareto distribution to data. *Journal of the American Statistical Association*, 92(440):1609–1620.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*, 7(2):174–196.
- Cui, W. and George, E. I. (2008). Empirical Bayes vs. fully Bayes variable selection. *Journal of Statistical Planning and Inference*, 138(4):888–900.
- Demirer, M., Diebold, F. X., Liu, L., and Yilmaz, K. (2017). Estimating global bank network connectedness. *Journal of Applied Econometrics*.
- Demos, A. and Sentana, E. (1998). An EM Algorithm for Conditionally Heteroscedastic Factor Models. *Journal of Business & Economic Statistics*, 16(3):357–361.
- Dempster, A. P., Laird, N. M., and Rubin, D. B. (1977). Maximum Likelihood from Incomplete Data via the EM Algorithm. *Journal of the Royal Statistical Society. Series B (Methodological)*, 39(1):1–38.

- Diebold, F. X. and Yilmaz, K. (2014). On the network topology of variance decompositions: Measuring the connectedness of financial firms. *Journal of Econometrics*, 182(1):119–134.
- Efron, B., Hastie, T., Johnstone, I., and Tibshirani, R. (2004). Least Angle Regression. *The Annals of Statistics*, 32(2):407–499.
- Elliott, M., Golub, B., and Jackson, M. O. (2014). Financial networks and contagion. *American Economic Review*, 104(10):3115–53.
- Embrechts, P., Klüppelberg, C., and Mikosch, T. (2013). *Modelling extremal events: for insurance and finance*, volume 33. Springer Science & Business Media.
- Engle, R. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *Journal of Economic Perspectives*, 15(4):157–168.
- Engle, R. F. (1982). Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation. *Econometrica*, 50(4):987–1007.
- Engle, R. F. and Bollerslev, T. (1986). Modelling the persistence of conditional variances. *Econometric Reviews*, 5(1):1–87.
- Fama, E. F. and French, K. R. (2015). A five-factor asset pricing model. *Journal of Financial Economics*, 116(1):1 – 22.
- Fan, J. and Li, R. (2001). Variable Selection via Nonconcave Penalized Likelihood and its Oracle Properties. *Journal of the American Statistical Association*, 96(456):1348–1360.
- Francq, C. and Zakoïan, J.-M. (2010). *GARCH models*. John Wiley & Sons, Ltd., Chichester.
- Friedman, J., Hastie, T., Höfling, H., and Tibshirani, R. (2007). Pathwise coordinate optimization. *The Annals of Applied Statistics*, 1(2):302–332.
- Friedman, J., Hastie, T., and Tibshirani, R. (2008). Sparse inverse covariance estimation with the graphical lasso. *Biostatistics (Oxford, England)*, 9(3):432–441.
- Friedman, J., Hastie, T., and Tibshirani, R. (2010). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software, Articles*, 33(1):1–22.
- Fusaro, P. (2005). Energy: an immature financial market. *Energy Hedge*, page 1.
- Gao, X. and Massam, H. (2015). Estimation of symmetry-constrained Gaussian graphical models: application to clustered dense networks. *Journal of Computational and Graphical Statistics*, 24(4):909–929.

- Geman, H. (2009). *Risk management in commodity markets: from shipping to agriculturals and energy*, volume 445. John Wiley & Sons.
- George, E. I. and McCulloch, R. E. (1993). Variable Selection Via Gibbs Sampling. *Journal of the American Statistical Association*, 88(423):881–889.
- George, E. I. and McCulloch, R. E. (1997). Approaches for Bayesian Variable Selection. *Statistica Sinica*, 7(2):339–373.
- Haldane, A. G. (2013). Rethinking the financial network. In *Fragile stabilität–stabile fragilität*, pages 243–278. Springer.
- Hansen, P. R. and Lunde, A. (2005). A forecast comparison of volatility models: does anything beat a GARCH (1, 1)? *Journal of Applied Econometrics*, 20(7):873–889.
- Hentschel, L. et al. (1995). All in the family: Nesting symmetric and asymmetric GARCH models. *Journal of Financial Economics*, 39(1):71–104.
- Hollo, D., Kremer, M., and Lo Duca, M. (2012). CISS - a composite indicator of systemic stress in the financial system. *ECB Working Paper Series*, 1426.
- Iori, G., De Masi, G., Precup, O. V., Gabbi, G., and Caldarelli, G. (2008). A network analysis of the italian overnight money market. *Journal of Economic Dynamics and Control*, 32(1):259–278.
- Jordan, I. K., Mariño-Ramírez, L., Wolf, Y. I., and Koonin, E. V. (2004). Conservation and co-evolution in the scale-free human gene coexpression network. *Molecular Biology and Evolution*, 21(11):2058–2070.
- Kliesen, K. L., Smith, D. C., et al. (2010). Measuring financial market stress. *Economic Synopses*.
- Korobilis, D. and Yilmaz, K. (2018). Measuring Dynamic Connectedness with Large Bayesian VAR Models. *Essex Finance Centre Working Papers*.
- Li, F. and Zhang, N. R. (2010). Bayesian variable selection in structured high-dimensional covariate spaces with applications in genomics. *Journal of the American Statistical Association*, 105(491):1202–1214.
- Liang, F., Paulo, R., Molina, G., Clyde, M. A., and Berger, J. O. (2008). Mixtures of g priors for Bayesian variable selection. *Journal of the American Statistical Association*, 103(481):410–423.

- Maruyama, Y. and George, E. I. (2011). Fully Bayes factors with a generalized g -prior. *The Annals of Statistics*, 39(5):2740–2765.
- McLachlan, G. and Krishnan, T. (2007). *The EM algorithm and extensions*, volume 382. John Wiley & Sons.
- McLachlan, G. and Peel, D. (2000). *Finite mixture models*. Wiley Series in Probability and Statistics: Applied Probability and Statistics. Wiley-Interscience, New York.
- Meinshausen, N. and Bühlmann, P. (2006). High-dimensional graphs and variable selection with the lasso. *The Annals of Statistics*, 34(3):1436–1462.
- Mitchell, T. J. and Beauchamp, J. J. (1988). Bayesian variable selection in linear regression. *Journal of the American Statistical Association*, 83(404):1023–1032.
- Müller, U. A., Dacorogna, M. M., Davé, R. D., Olsen, R. B., Pictet, O. V., and von Weizsäcker, J. E. (1997). Volatilities of different time resolutions-analyzing the dynamics of market components. *Journal of Empirical Finance*, 4(2-3):213–239.
- Nakatsuma, T. (1998). A Markov-chain sampling algorithm for GARCH models. *Studies in Nonlinear Dynamics & Econometrics*, 3(2).
- Nakatsuma, T. (2000). Bayesian analysis of ARMA-GARCH models: a Markov chain sampling approach. *Journal of Econometrics*, 95(1):57–69.
- Panagiotelis, A. and Smith, M. (2008). Bayesian identification, selection and estimation of semiparametric functions in high-dimensional additive models. *Journal of Econometrics*, 143(2):291–316.
- Park, T. and Casella, G. (2008). The Bayesian lasso. *Journal of the American Statistical Association*, 103(482):681–686.
- Pickands, J. (1975). Statistical inference using extreme order statistics. *The Annals of Statistics*, pages 119–131.
- Rizopoulos, D., Verbeke, G., and Lesaffre, E. (2009). Fully exponential Laplace approximations for the joint modelling of survival and longitudinal data. *Journal of the Royal Statistical Society. Series B. Statistical Methodology*, 71(3):637–654.
- Rothman, A. J., Levina, E., and Zhu, J. (2010). Sparse multivariate regression with covariance estimation. *Journal of Computational and Graphical Statistics*, 19(4):947–962.

- Ročková, V. and George, E. I. (2014). EMVS: The EM Approach to Bayesian Variable Selection. *Journal of the American Statistical Association*, 109(506):828–846.
- Scheipl, F., Fahrmeir, L., and Kneib, T. (2012). Spike-and-Slab Priors for Function Selection in Structured Additive Regression Models. *Journal of the American Statistical Association*, 107(500):1518–1532.
- Schweitzer, F., Fagiolo, G., Sornette, D., Vega-Redondo, F., Vespignani, A., and White, D. R. (2009). Economic networks: The new challenges. *Science*, 325(5939):422–425.
- Shumway, T. (1997). The delisting bias in CRSP data. *The Journal of Finance*, 52(1):327–340.
- Tibshirani, R. (1996). Regression Shrinkage and Selection via the Lasso. *Journal of the Royal Statistical Society, Series B*, 58:267–288.
- Yuan, M. and Lin, Y. (2006). Model selection and estimation in regression with grouped variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 68(1):49–67.
- Zhang, C.-H. (2010). Nearly unbiased variable selection under minimax concave penalty. *The Annals of Statistics*, 38(2):894–942.
- Zhao, T., Liu, H., and Zhang, T. (2018). Pathwise coordinate optimization for sparse learning: algorithm and theory. *The Annals of Statistics*, 46(1):180–218.
- Zou, H. (2006). The Adaptive Lasso and Its Oracle Properties. *Journal of the American Statistical Association*, 101(476):1418–1429.

A The SVSS–EM algorithm

We provide here a detailed proof of the updating equations of parameters $(\mu, \beta^\top, \varphi^\top, \nu_1^\top)^\top$ for the Heteroscedastic SSVS–EM algorithm introduced in Section 3. The E–step of the EM algorithm requires the evaluation of the conditional expectation of the complete–data log–likelihood in equation (19) of Section 3.2, leading to the EM objective function that is subsequently maximised to get parameters updates. Ignoring constant terms, the objective function of the EM algorithm becomes

$$\begin{aligned}
\mathcal{Q}(\boldsymbol{\vartheta} \mid \widehat{\boldsymbol{\vartheta}}^{(m)}) &= -\frac{T}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log(h_t) - \frac{1}{2} \sum_{t=1}^T \frac{\|y_t - \mu - \mathbf{x}_t^\top \beta - \mathbf{z}_t^\top \varphi\|_2^2}{h_t} \\
&\quad - \frac{p}{2} \log(2\pi) - \frac{1}{2} \sum_{j=1}^p \mathbb{E}_\gamma(\log(b_j)) - \frac{1}{2} \sum_{j=1}^p \beta_j^2 \mathbb{E}_\gamma\left(\frac{1}{b_j}\right) \\
&\quad + \sum_{j=1}^p \mathbb{E}_\gamma(\gamma_j) \log\left(\frac{\omega}{1-\omega}\right) + (p+b-1) \log(1-\omega) + (a-1) \log(\omega) \\
&\quad - (a+b+2) \sum_{j=1}^p \log(1+\nu_{1,j}) + b \sum_{j=1}^p \log(\nu_{1,j}) - \sum_{j=1}^p \log(\text{Be}(a+1, b+1)) \\
&\quad - \log(\text{Be}(a, b)) - \frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_\gamma^2) - \frac{1}{2} \frac{(\gamma_h - \mu_\gamma)^2}{\sigma_\gamma^2} \\
&\quad - \frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_\alpha^2) - \frac{1}{2} \frac{(\alpha_h - \mu_\alpha)^2}{\sigma_\alpha^2} \\
&\quad - \frac{1}{2} \log(2\pi) - \frac{1}{2} \log(\sigma_\beta^2) - \frac{1}{2} \frac{(\beta_h - \mu_\beta)^2}{\sigma_\beta^2} \\
&\quad - \frac{1}{2} (\varphi - \mu_\varphi)^\top \Sigma_\varphi^{-1} (\varphi - \mu_\varphi) \\
&= C + \mathcal{Q}_{1,c}(\mu, \beta, \varphi, \nu_1) + \mathcal{Q}_{2,c}(\nu_1) + \mathcal{Q}_{3,c}(\omega) + \mathcal{Q}_{4,c}(\psi) + \log \mathcal{Q}_{5,c}(\varphi),
\end{aligned}$$

where

$$\begin{aligned}
C &= -\frac{T+p+q+3}{2} \log(2\pi) - \log(\text{Be}(a+1, b+1)) \\
&\quad - \sum_{j=1}^p \log(\text{Be}(a, b)) - \frac{1}{2} \log |\Sigma_\varphi| \\
\mathcal{Q}_{1,c}(\mu, \beta, \varphi, \nu_1) &= -\frac{1}{2} \sum_{t=1}^T \log(h_t) - \frac{1}{2} \sum_{t=1}^T \frac{\|y_t - \mu - \mathbf{x}_t^\top \beta - \mathbf{z}_t^\top \varphi\|_2^2}{h_t} - \frac{1}{2} \sum_{j=1}^p \beta_j^2 \widehat{b_j^{-1}}^{(m)} \\
\mathcal{Q}_{2,c}(\nu_1) &= -\frac{1}{2} \sum_{j=1}^p \widehat{\log(b_j)}^{(m)} - (a+b+2) \sum_{j=1}^p \log(1 + \nu_{1,j}) + b \sum_{j=1}^p \log(\nu_{1,j}) \\
\mathcal{Q}_{3,c}(\omega) &= \sum_{j=1}^p \widehat{\gamma_j}^{(m)} \log\left(\frac{\omega}{1-\omega}\right) + (p+b-1) \log(1-\omega) + (a-1) \log(\omega) \\
\mathcal{Q}_{4,c}(\psi) &= -\frac{1}{2} \frac{(\gamma_h - \mu_\gamma)^2}{\sigma_\gamma^2} - \frac{1}{2} \frac{(\alpha_h - \mu_\alpha)^2}{\sigma_\alpha^2} - \frac{1}{2} \frac{(\beta_h - \mu_\beta)^2}{\sigma_\beta^2} \\
\mathcal{Q}_{5,c}(\varphi) &= -\frac{1}{2} (\varphi - \mu_\varphi)^\top \Sigma_\varphi^{-1} (\varphi - \mu_\varphi),
\end{aligned}$$

and

$$\widehat{\gamma_j}^{(m)} = \mathbb{E}_\gamma \left(\gamma_j \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, \widehat{\boldsymbol{\vartheta}}^{(m)} \right) = \frac{1}{1 + \widehat{d_j}^{(m)}} \quad (21)$$

$$\begin{aligned}
\widehat{b_j^{-1}}^{(m)} &= \mathbb{E}_\gamma \left(\frac{1}{b_j} \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, \widehat{\boldsymbol{\vartheta}}^{(m)} \right) \\
&= \frac{1}{\nu_{1,j}} \widehat{\gamma_j}^{(m)} + \frac{1}{\nu_0} \left(1 - \widehat{\gamma_j}^{(m)} \right)
\end{aligned} \quad (22)$$

$$\begin{aligned}
\widehat{\log(b_j)}^{(m)} &= \mathbb{E}_\gamma \left(\log(b_j) \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}, \widehat{\boldsymbol{\vartheta}}^{(m)} \right) \\
&= \log(\nu_{1,j}) \widehat{\gamma_j}^{(m)} + \log(\nu_0) \left(1 - \widehat{\gamma_j}^{(m)} \right),
\end{aligned} \quad (23)$$

for $j = 1, 2, \dots, p$, where $\widehat{d_j}^{(m)} = \sqrt{\frac{\widehat{\nu}_{1,j}^{(m)}}{\nu_0}} \exp \left\{ \frac{(\widehat{\beta_j}^{(m)})^2}{2} \left(\frac{1}{\widehat{\nu}_{1,j}^{(m)}} - \frac{1}{\nu_0} \right) \right\} \frac{1 - \widehat{\omega}^{(m)}}{\widehat{\omega}^{(m)}}$.

One nice feature of the EM algorithms would be the analytical solution of the M-step for all the parameters. Unfortunately, for the Gaussian regression model with Spike-and-Slab prior here considered, a closed form update does not exist for all the parameters. Therefore, we resort to a Conditional Expectation Maximisation (CEM) approach (see [McLachlan and Krishnan, 2007](#)). Under the Spike-and-Slab prior in equations (10)–(18), an iteration of the

CEM algorithm cycles through the set of all the model parameters. Specifically, we get

$$\hat{\boldsymbol{\vartheta}}_{\star}^{(m+1)} = \boldsymbol{\mu}_{\vartheta_{\star}} + \hat{\mathbf{K}}_{\star}^{(m)} (\mathbf{y} - \mathbf{X}_{\star} \boldsymbol{\mu}_{\vartheta_{\star}}) \quad (24)$$

$$\hat{\mathbf{K}}_{\star}^{(m)} = \hat{\boldsymbol{\Sigma}}_{\vartheta_{\star}}^{(m)} \mathbf{X}_{\star}^{\top} \hat{\mathbf{F}}^{(m)}{}^{-1} \quad (25)$$

$$\hat{\mathbf{F}}^{(m)} = \hat{\mathbf{H}}^{(m)}{}^{-1} + \mathbf{X}_{\star} \hat{\boldsymbol{\Sigma}}_{\vartheta_{\star}}^{(m)} \mathbf{X}_{\star}^{\top} \quad (26)$$

$$\hat{\omega}^{(m+1)} = \frac{1}{p+b+a-2} \left(\sum_{j=1}^p \hat{\gamma}_j^{(m)} + a - 1 \right) \quad (27)$$

$$\hat{\nu}_1^{(m+1)} = - \frac{(\hat{A}^{(m)} + \hat{B}^{(m)} + b) + \hat{\Delta}^{(m)}}{2(\hat{B}^{(m)} - a - 2)} \quad (28)$$

$$\hat{\Delta}^{(m)} = \sqrt{(\hat{A}^{(m)} + \hat{B}^{(m)} + b)^2 - 4(\hat{B}^{(m)} - a - 2)\hat{A}^{(m)}} \quad (29)$$

$$\hat{B}^{(m)} = -\frac{\hat{\gamma}^{(m)}}{2} \quad (30)$$

$$\hat{A}^{(m)} = \frac{1}{2} \hat{\boldsymbol{\beta}}^{(m)2} \odot \hat{\gamma}^{(m)}, \quad (31)$$

where $\boldsymbol{\vartheta}_{\star} = (\mu, \boldsymbol{\beta}^{\top}, \boldsymbol{\varphi}^{\top})^{\top}$, $\mathbf{y} = (y_1, y_2, \dots, y_T)^{\top}$, $\mathbf{X}_{\star} = \begin{pmatrix} \boldsymbol{\iota}_T & \mathbf{X} & \mathbf{Z} \end{pmatrix}$, $\boldsymbol{\iota}_T$ is a column vector of dimension $(T \times 1)$ of unit elements, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)^{\top}$ and $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_T)^{\top}$, $\boldsymbol{\mu}_{\vartheta_{\star}} = (\mu_{\mu}, \mathbf{0}^{\top}, \boldsymbol{\mu}_{\varphi}^{\top})^{\top}$, $\hat{\boldsymbol{\Sigma}}_{\vartheta_{\star}}^{(m)} = \text{diag}(\sigma_{\mu}^2, \hat{\mathbf{D}}^{(m)}, \boldsymbol{\Sigma}_{\varphi})$, \odot denotes the Hadamart componentwise vector multiplication, $\hat{\mathbf{D}}^{(m)} = \text{diag}\left\{\widehat{b_1^{-1}}^{(m)}, \widehat{b_2^{-1}}^{(m)}, \dots, \widehat{b_p^{-1}}^{(m)}\right\}$ and $\hat{\mathbf{H}}^{(m)} = \text{diag}\left\{\widehat{h_1}^{(m)}, \widehat{h_2}^{(m)}, \dots, \widehat{h_T}^{(m)}\right\}$. A detailed proof of the updating equations (24)–(26) is provided in Appendix A.1, while a detailed proof of the updating equations (28)–(31) is provided in Appendix A.2. Moreover, the updating equation (26) depends on the update of the matrix $\hat{\mathbf{H}}^{(m)}$ of variance terms. The update of the GARCH parameters is described in Appendix A.3.

A.1 Update of ϑ_*

Let $\mathcal{Q}_{\vartheta_*,c}(\mu, \beta, \varphi) = \mathcal{Q}_{6,c}(\mu, \beta, \varphi) + \mathcal{Q}_{7,c}(\varphi) + \mathcal{Q}_{8,c}(\mu)$ be the objective function with respect to the parameters $\vartheta_* = (\mu, \beta^\top, \varphi^\top)^\top$, where

$$\begin{aligned}\mathcal{Q}_{6,c}(\mu, \beta, \varphi) &= -\frac{1}{2} \sum_{t=1}^T \frac{\|y_t - \mu - \mathbf{x}_t^\top \beta - \mathbf{z}_t^\top \varphi\|_2^2}{h_t} - \frac{1}{2} \sum_{j=1}^p \beta_j^2 \widehat{b_j^{-1}}^{(m)} \\ \mathcal{Q}_{7,c}(\varphi) &= -\frac{1}{2} (\varphi - \mu_\varphi)^\top \Sigma_\varphi^{-1} (\varphi - \mu_\varphi) \\ \mathcal{Q}_{8,c}(\mu) &= -\frac{1}{2} \frac{(\mu - \mu_\mu)^2}{\sigma_\mu^2}.\end{aligned}$$

Upon defining the quantities $\mathbf{y} = (y_1, y_2, \dots, y_T)^\top$, $\mathbf{X}_* = \begin{pmatrix} \iota_T & \mathbf{X} & \mathbf{Z} \end{pmatrix}$, where ι_T is a column vector of dimension $(T \times 1)$, $\mathbf{X} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_T)^\top$ and $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_T)^\top$, $\mu_{\vartheta_*} = (\mu_\mu, \mathbf{0}^\top, \mu_\varphi^\top)^\top$, $\Sigma_{\vartheta_*} = \text{diag}(\sigma_\mu^2, \Sigma_\beta, \Sigma_\varphi)$ the objective function $\mathcal{Q}_{\vartheta_*,c}(\mu, \beta, \varphi)$ can be rewritten as

$$\begin{aligned}\mathcal{Q}_{\vartheta_*,c}(\mu, \beta, \varphi) &= -\frac{1}{2} (\mathbf{y} - \mathbf{X}_* \vartheta_*)^\top \mathbf{H}^{(m)-1} (\mathbf{y} - \mathbf{X}_* \vartheta_*) \\ &\quad - \frac{1}{2} (\vartheta_* - \mu_{\vartheta_*})^\top \Sigma_{\vartheta_*}^{-1} (\vartheta_* - \mu_{\vartheta_*}),\end{aligned}\tag{32}$$

thereby leading to the following first order condition with respect to the vector of unknown parameters ϑ_*

$$\frac{\partial \mathcal{Q}_{\vartheta_*,c}(\mu, \beta, \varphi)}{\partial \vartheta_*} = -\mathbf{X}_*^\top \mathbf{H}^{(m)-1} \mathbf{X}_* \vartheta_* + \mathbf{X}_*^\top \mathbf{H}^{(m)-1} \mathbf{y} - \Sigma_{\vartheta_*}^{-1} \vartheta_* + \Sigma_{\vartheta_*}^{-1} \mu_{\vartheta_*} = 0,\tag{33}$$

and to the parameter update $\widehat{\vartheta}_*^{(m+1)} = \left(\mathbf{X}_*^\top \mathbf{H}^{(m)-1} \mathbf{X}_* + \Sigma_{\vartheta_*}^{-1} \right)^{-1} \left(\mathbf{X}_*^\top \mathbf{H}^{(m)-1} \mathbf{y} + \Sigma_{\vartheta_*}^{-1} \mu_{\vartheta_*} \right)$. After some manipulations, we get

$$\begin{aligned}\left(\mathbf{X}_*^\top \mathbf{H}^{(m)-1} \mathbf{X}_* + \Sigma_{\vartheta_*}^{-1} \right)^{-1} &= \Sigma_{\vartheta_*} - \Sigma_{\vartheta_*} \mathbf{X}_*^\top \mathbf{F}^{-1} \mathbf{X}_* \Sigma_{\vartheta_*} \\ &= \Sigma_{\vartheta_*} - \mathbf{K}_* \mathbf{X}_* \Sigma_{\vartheta_*} \\ &= (\mathbf{I} - \mathbf{K}_* \mathbf{X}_*) \Sigma_{\vartheta_*} \\ \mathbf{K}_* &= \Sigma_{\vartheta_*} \mathbf{X}_*^\top \mathbf{F}^{-1} \\ \mathbf{F} &= \mathbf{H}^{(m)-1} + \mathbf{X}_* \Sigma_{\vartheta_*} \mathbf{X}_*^\top,\end{aligned}\tag{34}$$

and the parameters update $\hat{\boldsymbol{\vartheta}}_\star^{(m+1)}$ becomes

$$\begin{aligned}\hat{\boldsymbol{\vartheta}}_\star^{(m+1)} &= \boldsymbol{\mu}_{\vartheta_\star} - \boldsymbol{\Sigma}_{\vartheta_\star} \mathbf{X}_\star^\top \mathbf{F}^{-1} \mathbf{X}_\star \boldsymbol{\mu}_{\vartheta_\star} + \boldsymbol{\Sigma}_{\vartheta_\star} \mathbf{X}_\star^\top \mathbf{H}^{(m)-1} \mathbf{y} - \boldsymbol{\Sigma}_{\vartheta_\star} \mathbf{X}_\star^\top \mathbf{F}^{-1} \mathbf{X}_\star \boldsymbol{\Sigma}_{\vartheta_\star} \mathbf{X}_\star^\top \mathbf{H}^{(m)-1} \mathbf{y} \\ &= \boldsymbol{\mu}_{\vartheta_\star} - \mathbf{K}_\star \mathbf{X}_\star \boldsymbol{\mu}_{\vartheta_\star} + \boldsymbol{\Sigma}_{\vartheta_\star} \mathbf{X}_\star^\top \mathbf{H}^{(m)-1} \mathbf{y} - \mathbf{K}_\star \mathbf{X}_\star \boldsymbol{\Sigma}_{\vartheta_\star} \mathbf{X}_\star^\top \mathbf{H}^{(m)-1} \mathbf{y} \\ &= \boldsymbol{\mu}_{\vartheta_\star} + \mathbf{K}_\star (\mathbf{y} - \mathbf{X}_\star \boldsymbol{\mu}_{\vartheta_\star}),\end{aligned}\tag{35}$$

$$\tag{36}$$

where the last result follows immediately by observing that, from the definition of the Kalman gain $\mathbf{K}_\star \left(\mathbf{H}^{(m)-1} + \mathbf{X}_\star \boldsymbol{\Sigma}_{\vartheta_\star} \mathbf{X}_\star^\top \right) = \boldsymbol{\Sigma}_{\vartheta_\star} \mathbf{X}_\star^\top$, and

$$\boldsymbol{\Sigma}_{\vartheta_\star} \mathbf{X}_\star^\top \mathbf{H}^{(m)-1} \mathbf{y} = \mathbf{K}_\star \mathbf{H}^{(m)} \mathbf{H}^{(m)-1} \mathbf{y} + \mathbf{K}_\star \mathbf{X}_\star \boldsymbol{\Sigma}_{\vartheta_\star} \mathbf{X}_\star^\top \mathbf{H}^{(m)-1} \mathbf{y},\tag{37}$$

which is then substituted into equation (35), yields equation (36).

A.2 Update of ν_1

Concerning the solution for ν_1 observe that the update is the solution of the following equation with respect to ν_1 :

$$\hat{\nu}_1^{(m+1)} = \arg \max_{\nu_1 \in \mathbb{R}^+} \mathcal{Q}_{\nu_1},\tag{38}$$

where $\mathcal{Q}_{\nu_1} = \mathcal{Q}_{\nu_1,1} + \mathcal{Q}_{\nu_1,2} + \mathcal{Q}_{\nu_1,3}$

$$\mathcal{Q}_{\nu_1,j,1} = -\frac{1}{2} \widehat{\log(b_j)}^{(m+1)}\tag{39}$$

$$\mathcal{Q}_{\nu_1,j,2} = -\frac{1}{2} \hat{\beta}_j^{(m+1)2} \widehat{b_j^{-1}}^{(m+1)}\tag{40}$$

$$\mathcal{Q}_{\nu_1,j,3} = -(a + b + 2) \log(1 + \nu_{1,j}) + b \log(\nu_{1,j}).\tag{41}$$

Now, substituting for the expressions of $\widehat{\log(b_j)}^{(m+1)}$ and $\widehat{b_j^{-1}}^{(m+1)}$ into previous equations and differentiating with respect to ν_1 , the FOC of the optimisation problem becomes $\frac{\mathcal{Q}_{\nu_1}}{\partial \nu_1} = 0$, where

$$\frac{\mathcal{Q}_{\nu_1,j,1}}{\partial \nu_{1,j}} = -\frac{\hat{\gamma}_j^{(m+1)}}{2\nu_{1,j}} = 0\tag{42}$$

$$\frac{\mathcal{Q}_{\nu_1,j,2}}{\partial \nu_{1,j}} = \frac{\hat{\beta}_j^{(m+1)2} \hat{\gamma}_j^{(m+1)}}{2\nu_{1,j}^2} = 0\tag{43}$$

$$\frac{\mathcal{Q}_{\nu_1,j,3}}{\partial \nu_{1,j}} = -\frac{a + b + 2}{1 + \nu_{1,j}} + \frac{b}{\nu_{1,j}} = 0,\tag{44}$$

for $j = 1, 2, \dots, p$, leading to the following set of equations

$$\frac{A_j}{\nu_{1,j}^2} - \frac{a+b+2}{1+\nu_{1,j}} + \frac{B_j+b}{\nu_{1,j}} = 0 \quad (45)$$

$$-(a+b+2)\nu_{1,j}^2 + (B_j+b)\nu_{1,j}(1+\nu_{1,j}) + A_j(1+\nu_{1,j}) = 0 \quad (46)$$

$$(B_j-a-2)\nu_{1,j}^2 + (A_j+B_j+b)\nu_{1,j} + A_j = 0. \quad (47)$$

Therefore we have that

$$\hat{\nu}_{1,j}^{(m+1)} = \frac{-(A_j+B_j+b) - \Delta_j}{2(B_j-a-2)} \quad (48)$$

$$\Delta_j = \sqrt{(A_j+B_j+b)^2 - 4(B_j-a-2)A_j} \quad (49)$$

where $B_j = -\frac{\hat{\gamma}_j^{(m)}}{2} < 0$ and $A_j = \frac{1}{2}\hat{\beta}_j^{(m+1)2}\hat{\gamma}_j^{(m)} > 0$ for $j = 1, 2, \dots, p$ and $\hat{\nu}_1^{(m+1)} = (\hat{\nu}_{1,1}^{(m+1)}, \hat{\nu}_{1,2}^{(m+1)}, \dots, \hat{\nu}_{1,p}^{(m+1)})^\top$, $B = (B_1, B_2, \dots, B_p)^\top$ and $A = (A_1, A_2, \dots, A_p)^\top$, defined in Equations (30)–(31).

A.3 Updates of the GARCH parameters

The GARCH parameters $\psi = (\gamma_h, \alpha_h, \beta_h)$ do not admit a known closed form expression for the maximum of the log-likelihood because of the recursive nature of the variance equation in model (7). Therefore, we propose an approximate EM algorithm where the full conditionals of $(\gamma_h, \alpha_h, \beta_h)$ are approximated by a simpler model as in [Ardia \(2008\)](#) and [Nakatsuma \(1998, 2000\)](#) that introduced a similar approach in order to tailor the Metropolis–Hastings (MH) proposal to the target.¹¹ The approximating densities for the GARCH parameters are obtained by exploiting the well known ARMA representation of GARCH processes for $\varepsilon_t^2 = y_t - \mu - \mathbf{x}_t^\top \beta - \mathbf{z}_t^\top \varphi$, for $t = 1, 2, \dots, T$. Indeed, by defining the martingale difference process $v_t = \varepsilon_t^2 - \mathbb{E}_{t-1}(\varepsilon_t^2) = \left(\frac{\varepsilon_t^2}{h_t} - 1\right)h_t \sim (\chi_1^2 - 1)h_t$ with $\mathbb{E}(v_t | h_t) = 0$ and $\mathbb{E}(v_t^2 | h_t) = 2h_t^2$, we get the following stationary ARMA representation for the squared innovations ε_t^2

$$\varepsilon_t^2 = \gamma_h + (\alpha_h + \beta_h)\varepsilon_{t-1}^2 - \beta_h v_{t-1} + v_t, \quad (50)$$

¹¹As a further reference for an approximate EM within the context of multivariate GARCH models, see [Demos and Sentana \(1998\)](#).

for $t = 1, 2, \dots, T$, with $\varepsilon_0^2 = 0$ and $v_0 = 0$. Following [Ardia \(2008\)](#), expression (50) can be written as $v_t = \varepsilon_t^2 - \ell_t^{*\top} \tilde{\psi}_h$ where $\ell_t^* = (l_t^*, v_t^*)^\top$ and $\tilde{\psi}_h = (\gamma_h, \alpha_h)^\top$, with

$$l_t^* = 1 + \beta_h l_t^* \quad (51)$$

$$v_t^* = v_{t-1} + \beta_h v_{t-1}^*, \quad (52)$$

where the recursion is initialised at $l_0^* = v_0^* = v_0 = 0$. The function v_t in equation (50) can therefore be expressed as a linear function of the vector of transformed parameters $\tilde{\psi}_h$. The procedure then approximates the innovation term v_t in equation (50) by a Gaussian distribution $z_t \sim \mathbf{N}(0, 2h_t^2)$ leading to the following auxiliary model

$$z_t(\psi) = \varepsilon_t^2 - \gamma_h - (\alpha_h + \beta_h) \varepsilon_{t-1}^2 + \beta_h z_{t-1}(\psi), \quad (53)$$

where $z_t(\psi) = \varepsilon_t^2 - \ell_t^{*\top} \tilde{\psi}_h$, $\mathbf{z} = \boldsymbol{\varepsilon}^2 - \mathbf{L} \tilde{\psi}_h$, $\mathbf{L} = (\ell_1^\top, \ell_2^\top, \dots, \ell_T^\top)^\top$ and $\boldsymbol{\varepsilon}^2 = (\varepsilon_1^2, \varepsilon_2^2, \dots, \varepsilon_T^2)^\top$. Therefore, we can approximate the full conditional distribution of the GARCH parameters $\tilde{\psi}_h$ using the following the auxiliary model

$$\mathcal{L}^*(\gamma_h, \alpha_h \mid \mathbf{y}, \mathbf{X}, \mathbf{Z}) \propto |\boldsymbol{\Lambda}|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} \mathbf{z}^\top \boldsymbol{\Lambda}^{-1} \mathbf{z} \right\} \mathbb{1}_{\mathcal{S}}(\tilde{\psi}_h), \quad (54)$$

where $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}(\psi) = \text{diag}(\{2h_t^2(\psi)\}_{t=1}^T)$ and \mathcal{S} denotes the convex region such that $\gamma_h \geq 0$, $\alpha_h > 0$. The approximating density for γ_h, α_h is obtained by combining the approximated likelihood function defined in equation (54) and the prior density defined in equations (17)–(18), by the usual Bayesian updating

$$\pi(\tilde{\psi}_h \mid \beta_h, \mathbf{y}, \mathbf{X}, \mathbf{Z}) \propto \phi(\tilde{\psi}_h \mid \hat{\boldsymbol{\mu}}_{\tilde{\psi}_h}, \hat{\boldsymbol{\Sigma}}_{\tilde{\psi}_h}) \mathbb{1}_{(\mathcal{S})}(\tilde{\psi}_h), \quad (55)$$

where

$$\hat{\boldsymbol{\Sigma}}_{\tilde{\psi}_h} = (\mathbf{L}^\top \boldsymbol{\Lambda}^{-1} \mathbf{L} + \boldsymbol{\Sigma}_{\tilde{\psi}_h}^{-1})^{-1} \quad (56)$$

$$\hat{\boldsymbol{\mu}}_{\tilde{\psi}_h} = \hat{\boldsymbol{\Sigma}}_{\tilde{\psi}_h} (\mathbf{L}^\top \boldsymbol{\Lambda}^{-1} \boldsymbol{\varepsilon} + \boldsymbol{\Sigma}_{\tilde{\psi}_h}^{-1} \boldsymbol{\mu}_{\tilde{\psi}_h}), \quad (57)$$

and $\boldsymbol{\mu}_{\tilde{\psi}_h} = (\mu_\gamma, \mu_\alpha)^\top$ and $\boldsymbol{\Sigma}_{\tilde{\psi}_h} = \begin{pmatrix} \sigma_\alpha^2 & 0 \\ 0 & \sigma_\beta^2 \end{pmatrix}$ are the prior parameters defined in equations (17)–(18).

Concerning the remaining GARCH parameter β_h , the way it enters the function $z_t(\psi)$ in equation (53) prevents that $z_t(\psi)$ could be expressed as a linear function of β_h in a similar way as for $\tilde{\psi}_h$. To overcome this problem, we linearise $z_t(\psi)$ with respect to β_h by a first order Taylor

expansion at point $\hat{\beta}_h^{(m)}$

$$z_t(\beta) \approx z_t(\hat{\beta}_h^{(m)}) - \nabla_t^{\hat{\beta}_h^{(m)}} (\beta - \hat{\beta}_h^{(m)}), \quad (58)$$

where $\nabla_t^{\hat{\beta}_h^{(m)}} = -\frac{\partial z_t(\beta)}{\partial \beta} \Big|_{\beta=\hat{\beta}_h^{(m)}}$ where $\hat{\beta}_h^{(m)}$ is the value of the parameter β_h in the previous iteration of the EM algorithm and the term $\nabla_t^{\hat{\beta}_h^{(m)}}$ can be calculated by the following recursion

$$\nabla_t^{\hat{\beta}_h^{(m)}} = \varepsilon_{t-1}^2 - z_{t-1}(\hat{\beta}_h^{(m)}) + \hat{\beta}_h^{(m)} \nabla_{t-1}^{\hat{\beta}_h^{(m)}}, \quad (59)$$

with $\nabla_0^{\hat{\beta}_h^{(m)}} = 0$. Furthermore, let $r_t(\hat{\beta}_h^{(m)}) = z_t(\hat{\beta}_h^{(m)}) + \hat{\beta}_h^{(m)} \nabla_t^{\hat{\beta}_h^{(m)}}$ and $z_t(\hat{\beta}_h^{(m)}) = r_t(\hat{\beta}_h^{(m)}) - \hat{\beta}_h^{(m)} \nabla_t^{\hat{\beta}_h^{(m)}}$, then the main term in the quadratic form in equation (54) can be expressed as $\mathbf{z} = \mathbf{r} - \beta_h \nabla_\beta$, where $\nabla_\beta = (\nabla_1^{\hat{\beta}_h^{(m)}}, \nabla_2^{\hat{\beta}_h^{(m)}}, \dots, \nabla_T^{\hat{\beta}_h^{(m)}})^\top$ and $\mathbf{r} = (r_1(\hat{\beta}_h^{(m)}), r_2(\hat{\beta}_h^{(m)}), \dots, r_T(\hat{\beta}_h^{(m)}))^\top$ yielding the following approximation for the likelihood function of β_h

$$\mathcal{L}^*(\beta_h \mid \mathbf{Y}, \mathbf{X}, \mathbf{Z}, \gamma_h, \alpha_h) \propto |\Lambda|^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (\mathbf{r} - \beta_h \nabla_\beta)^\top \Lambda^{-1} (\mathbf{r} - \beta_h \nabla_\beta) \right\} \mathbb{1}_{(0, 1-\alpha_h)}(\beta_h). \quad (60)$$

The approximated full conditional distribution of the parameter β_h is obtained by combining the approximated likelihood function defined in equation (60) and the prior density defined in equations (17)–(18), by the usual Bayesian updating

$$\pi(\beta_h \mid \omega_h, \alpha_h, \hat{\beta}_h^{(m)}, \mathbf{Y}, \mathbf{X}, \mathbf{Z}) \propto \phi(\beta_h \mid \hat{\mu}_\beta, \hat{\sigma}_\beta^2) \mathbb{1}_{(0, 1-\alpha_h)}(\beta_h), \quad (61)$$

where

$$\hat{\sigma}_\beta^2 = \left(\nabla_\beta^\top \Lambda^{-1} \nabla_\beta + \frac{1}{\sigma_{\beta_h}^2} \right)^{-1} \quad (62)$$

$$\hat{\mu}_\beta = \hat{\sigma}_\beta^2 \left(\nabla_\beta^\top \Lambda^{-1} \mathbf{r} + \frac{\mu_{\beta_h}}{\sigma_{\beta_h}^2} \right). \quad (63)$$

Summarising, the updated values of the GARCH parameters at iteration $m+1$ are

$$\hat{\psi}_h^{(m+1)} = \left(\mathbf{L}^\top \Lambda^{-1} \mathbf{L} + \Sigma_{\hat{\psi}_h}^{-1} \right)^{-1} \left(\mathbf{L}^\top \Lambda^{-1} \boldsymbol{\varepsilon} + \Sigma_{\hat{\psi}_h}^{-1} \boldsymbol{\mu}_{\hat{\psi}_h} \right) \quad (64)$$

$$\hat{\beta}_h^{(m+1)} = \left(\nabla_\beta^\top \Lambda^{-1} \nabla_\beta + \frac{1}{\sigma_{\beta_h}^2} \right)^{-1} \left(\nabla_\beta^\top \Lambda^{-1} \mathbf{r} + \frac{\mu_{\beta_h}}{\sigma_{\beta_h}^2} \right), \quad (65)$$

where $\widehat{\boldsymbol{\psi}}_h^{(m+1)} = \left(\widehat{\gamma}_h^{(m+1)}, \widehat{\alpha}_h^{(m+1)} \right)^\top$ and

$$\mathbf{L} = (\ell_1^\top, \ell_2^\top, \dots, \ell_T^\top)^\top \quad (66)$$

$$\mathbf{\Lambda} = \mathbf{\Lambda}(\psi) = \text{diag} \left(\left\{ 2h_t^2 \left(\widehat{\psi}^{(m)} \right) \right\}_{t=1}^T \right) \quad (67)$$

$$\mathbf{r} = \left(r_1 \left(\widehat{\beta}_h^{(m)} \right), r_2 \left(\widehat{\beta}_h^{(m)} \right), \dots, r_T \left(\widehat{\beta}_h^{(m)} \right) \right)^\top \quad (68)$$

$$\nabla_\beta = \left(\nabla_1^{\widehat{\beta}_h^{(m)}}, \nabla_2^{\widehat{\beta}_h^{(m)}}, \dots, \nabla_T^{\widehat{\beta}_h^{(m)}} \right)^\top, \quad (69)$$

with $\ell_t^* = (l_t^*, v_t^*)$ defined in equations (51)–(52).

B Number of financial firms by country

The following Table provides the list of the 1248 financial firms considered in the empirical application discussed in Section 4.

Country	BAN	BRO	INS	OTH	Total	Country	BAN	BRO	INS	OTH	Total
ARGENTINA	6	0	0	3	9	LIECHT.	1	0	0	0	1
AUSTRALIA	4	6	6	15	31	LITHUANIA	1	0	0	1	2
AUSTRIA	4	1	2	2	9	LUXEMBOURG	1	1	0	7	9
BAHRAIN	6	2	1	3	12	MALAYSIA	7	0	2	9	18
BELGIUM	2	1	2	17	22	MALTA	2	0	1	3	6
BERMUDA	0	2	7	1	10	MEXICO	4	4	0	7	15
BRAZIL	4	2	2	5	13	MONACO	1	0	0	1	2
BULGARIA	5	0	0	8	13	MOROCCO	7	0	4	4	15
CANADA	8	2	7	17	34	NETHERLANDS	2	5	3	8	18
CHILE	6	2	2	4	14	NEW ZEALAND	1	0	0	7	8
CHINA	2	6	3	3	14	NIGERIA	11	0	1	2	14
COLOMBIA	7	3	0	1	11	NORWAY	4	1	3	1	9
CROATIA	3	0	2	1	6	OMAN	6	0	1	1	8
CYPRUS	2	2	0	9	13	PAKISTAN	4	0	1	0	5
CZECH R.	1	1	0	1	3	PERU	5	1	3	2	11
DENMARK	3	0	3	2	8	PHILIPPINES	7	0	0	6	13
EGYPT	7	2	0	4	13	POLAND	9	0	1	2	12
ESTONIA	0	0	0	3	3	PORTUGAL	2	0	0	2	4
FINLAND	2	0	1	3	6	QATAR	5	1	3	5	14
FRANCE	2	5	5	17	29	ROMANIA	1	1	0	4	6
GERMANY	6	1	3	18	28	RUSSIA	3	1	1	1	6
GREECE	3	1	2	4	10	SINGAPORE	1	2	1	23	27
GUERNSEY	0	0	0	9	9	SLOVAKIA	2	1	0	1	4
HONG KONG	3	3	2	22	30	SLOVENIA	0	1	3	3	7
HUNGARY	1	3	1	3	8	SOUTH AFRICA	5	1	5	5	16
INDIA	18	1	1	3	23	SPAIN	5	3	2	12	22
INDONESIA	6	0	1	4	11	SRI LANKA	7	0	0	2	9
IRELAND	1	1	1	2	5	SWEDEN	2	2	0	9	13
ISLE OF MAN	0	0	0	1	1	SWITZERLAND	10	6	7	12	35
ISRAEL	5	0	3	8	16	TAIWAN	5	2	3	1	11
ITALY	16	2	4	6	28	THAILAND	6	0	0	3	9
JAPAN	56	11	6	50	123	TURKEY	5	1	0	1	7
JERSEY	0	1	0	1	2	U.A.E	15	1	0	6	22
JORDAN	7	0	0	4	11	U.K.	10	13	11	90	124
KOREA S.	4	3	2	0	9	U.S.	32	18	29	63	142
KUWAIT	2	1	0	4	7	VENEZUELA	5	1	0	4	10

Table 5: List of the considered 1248 financial firms according to their sector type: banks “BAN”, brokers “BRO”, insurers “INS” and others “OTH”. The period is from 29 December 2000 to 6 October 2017 at a weekly frequency. The complete list of the firms is available upon request.

C Analysis on market returns with an additional set of variables

We provide an additional analysis to check if the Hubs indicator (HIND) maintains its explanatory power on the market returns where we perform a further regression by including three other sets of predictors lagged at one and four periods:

- Changes in volatility indices: VIX (U.S.), VSTOXX (EURO STOXX 50), VHSI (Hong Kong) and VNKY (Nikkei). Results are included in Table 6;
- Changes in financial stress indicators such as the CISS and the STLFSI (Table 7);
- The Fama and French five global factors (Fama and French, 2015): i) Mkt-Rf (the market's excess return); ii) SMB (Small Minus Big); iii) HML (High Minus Low); iv) RMW (Robust Minus Weak) and v) CMA (Conservative Minus Aggressive)¹². Results are reported in Table 8.

¹²The Fama and French five factors makes use of six value-weight portfolios formed on size and book-to-market, the six value-weight portfolios formed on size and operating profitability, and the six value-weight portfolios formed on size and investment. For further information, see http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_5_factors_2x3.html.

	estimate	std. err.	t-stat	p-value
<i>mean equation</i>				
(constant)	0.0008	0.0011	0.7200	0.4700
r_{t-1}	0.0282	0.0537	0.5300	0.5990
$\log(h_{t-1})$	0.5459	1.6743	0.3300	0.7440
ΔHIND_{t-1}	-0.1277	0.0346	-3.6900	0.0000
ΔHIND_{t-4}	0.1042	0.0489	2.1300	0.0330
ΔVIX_{t-1}	-0.0006	0.0006	-1.0000	0.3150
ΔVIX_{t-4}	-0.0011	0.0005	-2.0500	0.0400
$\Delta\text{VSTOXX}_{t-1}$	0.0009	0.0005	1.8100	0.0710
$\Delta\text{VSTOXX}_{t-4}$	0.0003	0.0005	0.6400	0.5210
ΔVHSI_{t-1}	0.0005	0.0004	1.0800	0.2800
ΔVHSI_{t-4}	-0.0001	0.0005	-0.2200	0.8290
ΔVNKY_{t-1}	0.0000	0.0004	0.1000	0.9170
ΔVNKY_{t-4}	0.0003	0.0003	1.0100	0.3120
<i>volatility equation</i>				
(constant)	-0.1645	0.0510	-3.2200	0.0010
z_{t-1}	-0.1418	0.0141	-10.0500	0.0000
$ z_{t-1} - E z_{t-1} $	0.0622	0.0236	2.6300	0.0080
$\log(h_{t-1})$	0.9784	0.0068	143.5200	0.0000
obs	767		Wald $\chi^2(12)$	31.33
Log-likelihood	1786.77		Prob > χ^2	0.0018

Table 6: Estimates for the AR(1)–EGARCH–m(1,1) model by including the (lagged) changes in the HUBS indicator ΔHIND_t and the changes in volatility indices: VIX CBOE (S&P 500), VSTOXX (EURO STOXX 50), VHSI (Hong Kong) and VNKY (Nikkei). The considered period is from 20 December 2002 to 6 October 2017.

	estimate	std. err.	t-stat	p-value
<i>mean equation</i>				
(constant)	0.0010	0.0010	1.0200	0.3090
r_{t-1}	-0.0277	0.0406	-0.6800	0.4950
$\log(h_{t-1})$	0.2719	1.5517	0.1800	0.8610
ΔHIND_{t-1}	-0.1329	0.0339	-3.9200	0.0000
ΔHIND_{t-4}	0.0965	0.0484	1.9900	0.0460
ΔCISS_{t-1}	0.0002	0.0227	0.0100	0.9940
ΔCISS_{t-4}	-0.0005	0.0228	-0.0200	0.9840
$\Delta\text{STLFSI}_{t-1}$	0.0024	0.0094	0.2600	0.7950
$\Delta\text{STLFSI}_{t-4}$	-0.0083	0.0112	-0.7400	0.4590
<i>volatility equation</i>				
(constant)	-0.1560	0.0490	-3.1900	0.0010
z_{t-1}	-0.1395	0.0152	-9.1700	0.0000
$ z_{t-1} - E z_{t-1} $	0.0582	0.0221	2.6300	0.0090
$\log(h_{t-1})$	0.9796	0.0065	149.6100	0.0000
obs	767		Wald $\chi^2(8)$	18.57
Log-likelihood	1781.07		Prob> χ^2	0.0173

Table 7: Estimates for the AR(1)–EGARCH–m(1,1) model by including the (lagged) changes in the Hubs indicator ΔHIND_t and the changes in financial stress indicators such as the CISS and the STLFSI. The considered period is from 20 December 2002 to 6 October 2017.

	estimate	std. err.	t-stat	p-value
<i>mean equation</i>				
(constant)	0.0008	0.0010	0.8200	0.4130
r_{t-1}	-0.1105	0.0931	-1.1900	0.2350
$\log(h_{t-1})$	0.4977	1.4337	0.3500	0.7280
ΔHIND_{t-1}	-0.1432	0.0339	-4.2300	0.0000
ΔHIND_{t-4}	0.1077	0.0487	2.2100	0.0270
Mkt-Rf $_{t-1}$	0.0672	0.1052	0.6400	0.5230
SMB $_{t-1}$	0.0216	0.1094	0.2000	0.8440
MHL $_{t-1}$	0.1028	0.1354	0.7600	0.4480
RMW $_{t-1}$	-0.2065	0.1883	-1.1000	0.2730
CMA $_{t-1}$	-0.0113	0.1911	-0.0600	0.9530
<i>volatility equation</i>				
(constant)	-0.1603	0.0528	-3.0300	0.0020
z_{t-1}	-0.1452	0.0153	-9.5000	0.0000
$ z_{t-1} - E z_{t-1} $	0.0507	0.0234	2.1700	0.0300
$\log(h_{t-1})$	0.9791	0.0071	138.6100	0.0000
obs	767		Wald $\chi^2(9)$	24.00
Log-likelihood	1782.43		Prob> χ^2	0.0043

Table 8: Estimates for the AR(1)–EGARCH–m(1,1) model by including the (lagged) changes in the Hubs indicator ΔHIND_t and the Fama and French five factors (Fama and French, 2015). The considered period is from 20 December 2002 to 6 October 2017.

Recent Issues

No. 243	Nicoletta Berardi, Marie Lalanne, Paul Seabright	Professional Networks and their Coevolution with Executive Careers: Evidence from North America and Europe
No. 242	Ester Faia, Vincenzo Pezone	Monetary Policy and the Cost of Wage Rigidity: Evidence from the Stock Market
No. 241	Martin Götz	Financial Constraints and Corporate Environmental Responsibility
No. 240	Irina Gemmo, Martin Götz	Life Insurance and Demographic Change: An Empirical Analysis of Surrender Decisions Based on Panel Data
No. 239	Paul Gortner, Baptiste Massenet	Macroprudential Policy in the Lab
No. 238	Joost Driessen, Theo E. Nijman, Zorka Simon	Much Ado About Nothing: A Study of Differential Pricing and Liquidity of Short and Long Term Bonds
No. 237	Nathanael Vellekoop	Explaining Intra-Monthly Consumption Patterns: The Timing of Income or the Timing of Consumption Commitments?
No. 236	Aleksey Kolokolov, Giulia Livieri, Davide Pirino	Statistical Inferences for Price Staleness
No. 235	Christian Kubitz, Lorian Pelizzon, Mila Getmansky Sherman	The Pitfalls of Central Clearing in the Presence of Systematic Risk
No. 234	Alejandro Bernales, Nicolás Garrido, Satchit Sagade, Marcela Valenzuela, Christian Westheide	A Tale of One Exchange and Two Order Books: Effects of Fragmentation in the Absence of Competition
No. 233	Baptiste Massenet, Yuri Pettinicchi	Can Households See into the Future? Survey Evidence from the Netherlands
No. 232	Jannic Alexander Cutura	Debt Holder Monitoring and Implicit Guarantees: Did the BRRD Improve Market Discipline?