Networks of spillover effects to spot systemic risk in the banking industry: testing Granger causality in a High Dimensional VAR model.

Lund University, School of Economics and Management

Department of Economics

Essay seminars: 29/*05 – 02/06/2023*

Author: Emanuele Paccagnella

Supervisor: Luca Margaritella

Table of contents:

Introduction

The recent turmoil in the banking industry, and the collapse of Silicon Valley Bank (SVB) alongside the crisis of many others, once again brought up the fear among the public of a new recession affecting the global economy. Moreover, as in recent history has been many times the case, more abruptly than ever in 2008, the downfalls of U.S. banks, have affected within a small time frame the banks and the financial institutions across the ocean. The reasons of these co-movements in banks wellbeing can be ascribed to many factors. First, firms carrying out businesses together or in the same sector are affected by the same risks (for instance same credit risks). Second, banks whose balance sheets are closely related suffer from the same kinds of macroeconomic variations (such as an increase in interest rates by the central bank as was the case for SVB). Third, bank-runs triggered on a cross-border level by the fall of a big financial institution tend to spread panic across the depositors, ultimately threatening to create a huge cascade effect. Hence it is of great public interest to understand how specific institutions health and performance will affect each other, much like in a network of infrastructures, that ultimately may highlight a deep systemic risk or a dangerous situation for regulators to deal with. Moreover it is often the case that trying to analyse each single bank's balance sheet is of great importance both for financial supervisors and central banks, in order to assess the activities of the firm and the risks the specific bank is subject to, it may however become really difficult to put these information together and assess the level of entanglement of the banking industry, and more importantly to understand the way specific banks behave when compared to one another, ultimately allowing governments and regulators to undertake timely and aimed actions to counteract the crisis spreading from one singular bank or a set of them.

In this thesis we use the econometric toolbox of estimation of Granger Causality relations in a high dimensional Vector Autoregressive (VAR) model. The objective is to shape a network of spillover effects among a pool of banks, to depict the degree of possible contagion in a financial distress moment. The VAR is specifically casted on a large sample of the banking industry with the purpose of determining whether it exists a significant dynamic effect in time of the performance of pairs of banks with the respect to the others. Or in other words we're interested in answering to the question "Does the degree of predictability of a bank performance increase when conditioning on performances previously recorded by other banks?”. The magnitude of this relationship will depend on the specific macroeconomic and industrial environment that banks pertaining to a certain geographic or economic area will be subject to. Furthermore, it is reasonable to expect the network of Granger causal relations to assume different degrees of connectivity over time. Therefore, we’ll highlight this behaviour in the network results and try to pair it with relevant macroeconomic events, ultimately providing a simple interpretation of the evolution over time.

However, because of the econometric testing procedure that we’re using, we’ll just be able to tell whether the causing relation is there or not, but not why it exists, it will take a further breaking down of the relationship (mainly by looking at specific balance sheets and activities) to understand the nature of the relation. The aim of the research is therefore to analyse the existence and the extent of such interconnections among banks, and ultimately their evolution over time.

The definition of Granger Causality will force us to account for all the possible information (control variables) at disposal, leading into a context of High Dimensionality (HD). The problems related to HD will be dealt with by using a variable selection method, the Post double selection proposed by Belloni et al. 2014, and expanded by Margaritella et al. 2021 (PDS-LM VAR), that rests on the notions of Lasso estimation and sparsity.

Methods:

Before proceeding with the analysis of our main topic some considerations of general term are to be drawn, together with a brief review of some relevant literature in the field.

The methods and indicators used in tracking systemic risk, or to alert regulators about a risk of financial distress are of various nature and they rest on different financial variables Tressel *et al.* (2022).

The stock price of a bank and its evolution over time, offer a good depiction of the general well being of a bank (in a way in financial literature the evolution of price is deemed to condense “all the informations” available to the public and to the market). Using the concept of Granger Causality we detect the degree of interdependence among individuals in the banking industry, which seems to be a good depiction of exposure to systemic risk.

Alternative methods: Sparse vector heterogeneous autoregressive modeling for realized volatility and long memory phenomena.

Baek and Park 2021 construct a model of Heterogeneous Autoregression to analyse realized volatilities in a multivariate case, the work rests on the pillars built in the field by *Andersen et al. 2003* and *Corsi 2009*, and stems from the growing necessity to model co-movements in network of interactions and spill over effects in a time series context. Realized volatility is a measure (investigated extensively by *Andersen et al. 2003)* of volatility that can be extracted from data and computed over a certain frequency. The framework is particularly suitable when dealing with portfolio and risk management, and, more generally, financial measures. The framework proposes to accurately predict the linkages and interactions occurring in a network (in *Baek and Park,* twenty financial indexes that show a certain level of financial integration) jointly, and in a multivariate context. These papers are of interest in shaping some issues related to the frequency of data (particularly in case of realized volatilities) that are interesting to understand even when dealing with level variables that are not subject to the same kind of problems. This literature also introduces issues of the long -memory phenomena that data may be subject to.

Volatility heterogeneous AR model (Corsi 2009): Long Memory

The paper proposes a model able to consistently estimate financial returns evolution over time, so suitable to our analysis of stock returns. Financial data of this kind are most of the times subject to problematics in their modelling, requiring specific modelling tools. Long memory volatility is obtained by employing fractional difference operators, meaning heterogeneous lags: such as using in the same autoregressive model a day lag, a week lag and a month lag at the same time; in order to have comparable measures of realized volatility over different time horizons. However, the method of fractional difference filter is critiqued in lacking clear economic interpretation, instead only being able to accurately predict the long-memory volatility effects in the dataset.

There's a difference between actual long-memory volatility patterns and simple component model but these last ones are much easier to estimate, and since in empirical settings the two different things are hard to tell apart it may be just easier to model a simple component model.

The idea to model heterogeneity, as proposed by *Corsi 2009*, stems from the financial context and the nature of realized volatilities. The idea is as follows: different agents in the market trade on different time horizons and with different frequencies, such that certain traders manage their portfolios on an intraday horizon taking into account high frequency informations, and pension funds or other financial institutes trade on a much longer term horizon and with a lower frequency approach to informations. Then it is reasonable to expect the realized volatilities will be fractioned into different volatilities components contributing to the dependent variable were modelling. Generally, the idea can be summarized by the quote "the multicomponent structure stems from the heterogeneous nature of the information arrivals", referring to the fact that each agent contributing to the volatility will be subject to some heterogeneity, in the specific circumstance of *Corsi* this is due to different frequency of information that the agents are associated to. Then it is reasonable to model variables in three chronological dimensions: the day (subject to operations of short-term agents), the week (subject to operations of medium-term agents) and the month (subject to long term ones). This setting also offers us a simple and appealing econometric interpretation; this allows to analyse in a more isolated context the dynamics of components of the network we're investigating. *Corsi* concludes upon further analysis that the volatility of long term has a stronger effect on short term volatility than it is the other way around ("volatility cascade from low frequencies to higher frequencies"). Conceptually this can be explained by the fact that short term traders take into account the expected level of long-term volatility, since this one determines the future level of risk and exposure to loss (short traders adapt to volatility in the long term by reassessing their current positions hence producing volatility in the short term). The HAR RV proposed by *Corsi* can be defined as a cascade model defined for different frequencies (different components) in the usual context of autoregressive model (it can be seen simply as a three-factor stochastic volatility model, where each factor corresponds to the past realized volatilities at different frequencies). This model is found to appropriately model the volatility memory, and to accurately represent the long memory of the empirical volatility (even though the HAR model itself is a model of short memory, model long memory volatility in a simple and parsimonious way, showing good forecasting performances). The big advantage of this model compared to ARFIMA is the simplicity and the easiness in extending it by adding new variables, to account for different components of heterogeneity.

The VHAR model produces three k*k matrices of parameters, representing: daily, weekly, and monthly relationships among elements of the analysis. It is likely that in this setting we find ourselves in a situation of high dimensions and low sample size, as such (similarly to our main situation) we won't be able to estimate the model using traditional estimation methods as OLS; instead we'll need to rely on sparsity seeking/variable selection methods as adaptive LASSO and standard LASSO (is later shown to poorly perform in this context, bringing with it a big bias issue. The paper finds that VHAR model improves realized volatilities forecasting. The sparse VHAR is compared against a non-sparse usual OLS model and a usual lasso (adaptive weight = 1). Note the advantages of multivariate VHAR model over the univariate HAR: the VHAR specifically quantifies the effect of one element (stock market index/bank) on all the others one by one; allowing for both a more precise estimation and prediction ability as provided by the analysis in the paper, and a punctual and clear representation of the network behaviours (by making it easier to outline the edges connecting different elements, or representing the relationships as a matrix of coefficients where the magnitude of each one can be highlighted and the most significant relations and individuals can be identified).

The heterogeneous modelling approach is however not considered further in this thesis. Also consider that the medium-term nature of the dataset used (17 years) seems too small to be subject to long memory effects. The issue may however arise in longer time spans considered.

The validity of Post Double Selection

In the main paper of reference to this work, *Margaritella et al. 2021* consider the realized volatilities of 30 stock returns, to identify network of volatility spillovers in a high dimensional VAR setting. Post Double Selection (PDS) procedure is proposed as a valid variable selection method to cope with the High Dimensionality originated. (The method itself is explained in detail in post double selection section). The analysis is conduced upon a big-time horizon, this allows to obtain a huge number of observations that allow to use a standard "full var" (without employing any variable selection and sparsity tool) alongside the PDS Var, to outline a comparison in performance and understand whether this double post selection method is worth the use. The results on a long-time horizon dataset employing the two different kinds of VAR model take into consideration realized variances of a variable (namely stock returns) and provide similar results. However, when considering a different time horizon the HD becomes an issue once again, and the standard full VAR model is not an available choice anymore. "Graph theory" is employed to identify cluster of elements which are closely related allowing to identify so called "communities" (densely related within group, and poorly related across groups). PDS-LM Var seems the best model in identifying such communities and shaping them. While the result obtained over the full time period using the full-var model is almost the same, when we move to the shorter time period, such introducing HD, the PDS-LM VAR model is the only able to identify the same community patterns identified over the whole time window (confirming the exogenous spill over effects existing among individuals and previously found). Note that correctly identify the spillover effects and the communities' pattern in the end is our main objective.

Increasing the information set and considering a high dimensional VAR instead of bivariate, *Margaritella et al.* are able to obtain more realistic effects than in two dimensions; BiHVAR (bivariate VAR testing) just poorly performs in both settings. After these considerations, and the performance comparison proposed in *Margaritella et al*., the PDS-LM VAR model seems the best way to deal with our proposed analysis of Granger Causal relations.

We won't model measures of realized volatilities, however. Instead, we'll deal with pure stock prices expressed in nominal terms.

Granger causality notions and Information set

The concept of Granger Causality can be defined as capturing predictability, *given (conditional on)* *a particular information set*. If the distribution of a variable Y conditional on an information set Ω which is made from all the information available in the world at time t-1, changes when removing the variable "X" at t-1 from Ω, then we can say X "Granger causes" Y. The term ''causality'' should be read as ''predictability'' in practice. Particularly in the context of time series two main features were identified as a ground to identify causal relationships, these are: temporal precedence, the causes should precede the effects, and physical influence, manipulations of the causes change the effects ("enforced changes in the causing variable will induce changes on the caused variable").

Eichler 2013 identifies different kinds of causality in the context of time series analysis: like intervention causality, stating that a cause-effect relationship should persist when the causing variable is manipulated, whereas any non-causal relation should vanish. Note however that: in time series, controlled experiments cannot be carried out, hence, effect of interventions can only be detected in the distribution. This effect can be detected only if it affects just the distribution of the variable of interest whereas all other conditional distributions remain the same.

Granger causality relies on the definition of a probabilistic causal relationship such that we're able to define it in empirical settings. As such GC takes more the shape of an increase in predictability that some information can add, than that of a proper causal relationship which can only be identified by means of train and control groups. The concept relies on two characteristics. The first is temporal precedence, as before this is the basis for all concepts of causality. The second is slightly more stringent and states that: the causal series should contain unique information about the caused variables that would otherwise be unavailable (such that the predictability of our interest variable is enhanced). The second element allowing to define granger causality, "unique information", first requires the definition of two information sets: the set containing all information in the universe up to time t, and the set containing all information in the universe aside from the information provided by X (the allegedly causal variable) up to time t. Then if Y is caused by X, we would expect the two distributions of Y conditional on the two different information sets to differ from one another. This kind of definition is however hard if not impossible to detect in actual data settings, to define the complete information set will be a hard task. However, the definition involving information sets clearly stresses the idea of Granger that the information set should be chosen as large as possible, so that all possibly relevant variables will be included. So, GC definition remains strongly tied to its identification always conditionally upon a certain information set, that we want to contain all the information in our interest and available to us (to be controlled for). Then the main task in an empirical context is that of the inclusion of all relevant variables, to satisfy the notion of complete information; the reason is, not including a variable that affects two or more of the observed variables (confounders) may turn out being wrongly interpreted as Granger causality results, even though just spurious causal. Spurious causalities can be understood conceptually in this way: the omission of relevant variables may induce associations between the lagged variable, that are then wrongly interpreted as causal relations, and that disappear if the relevant variables are added. Thus, in our VAR analysis as discussed, we should try to involve all the relevant variables.

<u>High dimensionality and GC testing</u>

To avoid spurious causality generated by a missing variable, one may want to include all the variables at disposal in the regression, however this may introduce a problem of high dimensionality, that becomes even more troublesome in the present VAR context, where the number of parameters increases quadratically with the number of series included in the system. This high dimensionality (specially in a context where T dimension is relatively short) prevents the estimation by standard means of Ordinary Least Squares (OLS) or (ML). Reason being that overfitting and high-variance estimates are obtained in high dimensional regression ("curse of dimensionality''); and, even more importantly, in cases the number of parameters is higher than the number of observations OLS is not feasible at all.

The Vector Autoregressive Model of order "p", VAR(p), identifies a k-dimensional vector Yt (all the individuals, or variables in a Var context), which is modelled as the autoregression with "p" lags:

$$Yt \;=\; A1\, Y_{t-1} \;+\; \ldots \;+\; Ap\, Y_{t-p} \;+\; ut$$

Where A1…Ap are k*k matrices of parameters (hence k^2*p total parameters) and ut is a disturbance vector that is a martingale difference sequence (MDS), which equals to saying its expected value conditionally on the "past" is zero (behaves as white noise).

The bivariate Granger causality can be tasted on the VAR model by individually testing the coefficients (via t-test or Wald test) and jointly significance testing all the elements "i,j" for every i = 1,…,I and j = 1,…,J; in the matrices A1,…, Ap (via F-test). We'll test every pair of individuals (banks) for GC relation for 87 x 87 possible relations. GC is considered a feature of the "population" hence a joint significance test of the coefficients is necessary. Also, note that more precisely we test for Granger "non-causality" from a group J to I, as we're testing $H_0 : A1 = A2 = \cdots = Ap = 0$, if we reject this null hypothesis, we'll find Granger Causality.

In a case where p = 1 we can easily understand the situation by defining a group J (Granger Causing), group I (Granger Caused); then the VAR model is defined as the distribution of $Y\ I, t$ conditional on $Y\ J, t-1$; then if the coefficient $A\ i, j$ = 0 we'll have no Granger Causality (this is for simplicity, but in reality we test conditionally on the whole information set Ω, HD). In most cases as in ours, J and I will only correspond to one variable (one bank). We test GC from one bank J to one bank I conditional on all the others.

The issue arising with high dimensionality is not only that of dealing with estimation but also the interpretation becomes a hard task for the researcher. *Belloni et al. 2014* discuss how controlling for so many variables may render almost impossible to really grasp the effect of the treatment variable that we're truly interested in. In this context we say exogeneity holds for the treatment variables (potentially p > n if High Dimensional) once we account for a certain number (s < n) of control variables. Then the problem of estimating the coefficients of the treatment variables (the ones we want to test for GC) becomes a variable selection problem. Assuming there exist many potentially "p" controls, of which a small minority "s" is sufficient to correctly approximate and represent the DGP. The true DGP is given by a linear combination of treatment and a function of control variables to be correctly identified, it is objective of the researcher to best approximate and identify the function of controls that characterizes DGP. Post double selection procedure proposed by *Belloni et al.*, later analysed (in section Post Double Selection procedure), performs estimation and inference on "structural effects" in an environment where treatment variables can then be considered as exogenous.

We want thus to impose on our coefficients a sparsity pattern, trying to reduce the number of covariates to allow for a much easier interpretation of the key explanatory variables. The idea of sparsity finds an intuition in the conceptual framework that is the following: in economics research many times we expect to be able to capture some complex phenomena by observing just a small number of relevant variables representative of the DGP.

In detecting Granger causality, the concept of sparsity will become a method to overcome the issue of High Dimensionality by identifying only the relevant control variables (individuals).

LASSO estimation

The most used estimation methods to deal with seeking/imposing sparsity are Ridge and Lasso regression. In this thesis we're interested on the Lasso method as it is the basis of Post Double Selection method, and it solves the issues arising in Ridge regression related to the variable selection *(Tibshirani 1996).* First, we assume sparsity on β vector, that is assuming β vector of coefficients can be correctly approximated by a vector containing a large (significant) portion of coefficients equal to 0. "The sparsity assumption validates the use of variable selection methods, thereby reducing the dimensionality of the system without having to sacrifice predictability" *Margaritella* et al. We then proceed to minimize a loss function, which is the usual residual sum of squares with the novelty being an "L1 norm constraint", a weighted penalty term that is governed by the magnitude of the Beta coefficient and a tuning parameter "λ ". For a general n-dimensional vector of responses y and n × M-dimensional matrix of covariates X, the lasso simultaneously performs variable selection and estimation of the parameters by solving:

$$\hat{\beta}(\lambda) = \arg\min_{\beta}\left(\frac{1}{T}\,\|y - X\beta\|_2{}^2 \quad + \lambda \sum_{m=1}^{M} |\beta m| \right)$$

where λ is a non-negative tuning parameter determining the strength of the penalty. The notation βˆ(λ) highlights that the solution to the minimization problem depends on λ, the choice of which is not unique: in the context of Var models*, Margaritella et al*. compares different methods for the selection of the tuning parameter (various methods exist to choose λ, the ones found to best perform in our context consists in using Bayesian Information Criteria). Also, it can be noted that the number of selected coefficients is a function of λ; the L1 norm constraint has the property of constraining the sum of all the beta coefficients, such that some of them must be forced equal to zero by the estimation process. Lasso succeeds in performing variable selection and reducing the effect of overfitting *(Tibshirani).*

The use of Lasso to detect Granger Causality in High Dimensionality

Oracle property is intended as the ability to correctly detect the sparsity pattern, by correctly identifying the zero parameters as being zero and the non-zero parameters as being non-zero. *Kock And Callot 2015* prove the asymptotic efficiency and oracle ability of LASSO and Adaptive LASSO in a context of sparsity in DGP; with the advantage of LASSO being feasible even with fewer observations than variables (HD) and that means also in our setting HD-VAR for granger causality. The validity of these methods also allows to focus on shorter and most recent time spans than being obliged to go back huge time spans to obtain more observations. Since adaptive LASSO can be used to correctly detect zero and non-zero parameters this can be used in testing for HD Granger Causality. (Adaptive lasso is asymptotically oracle-efficient as the OLS, probability of selecting the correct sparsity pattern tending to 1, lasso asymptotically and non-asymptotically correctly identifying zero and non-zero parameters). However, we'll use LASSO in a different way than to directly test for granger Causality.

In our empirical setting we'll use Lasso to perform control variable selection. We'll then be able in the linear estimation post selection to detect a Granger Causality relation going from the treatment to the outcome variable in an exogenous setting, as required by the notion of Granger Causality and as discussed in Eichler et al. Using just one step of Lasso selection, however, is not a feasible option.

The Post Double selection method: procedure and overview

An issue that arises often when trying to inference on causal effects is to identify which control variables allow to perform inference in a randomic (population like) setting, otherwise meeting the risk of wrongly detecting spurious relationships or reaching biased conclusions. The randomic setting can be obtained as a set of control variables of interest can be taken to be randomly assigned if we are controlling for some variables in the population. Most of the time this will be coped with by identifying these control variables from the economic theory, or report results after using different set of control variables for each one of them; these methods however are not automatically the best to be used.

Even though, as said, we employ Lasso for the selection of controls, LASSO on its own creates problems when dealing with "post-selection inference", the LASSO "oracle use" such as adaptive lasso, produces what we may consider an omitted-variable bias, ruling out variables which in a common estimation would otherwise be significant. When in presence of coefficients close but not equal to zero, Lasso may wind up discarding potentially important covariates, not being able to distinguish them as being qual to zero or not. This can be solved by imposing some selection criteria and conditions to rule out small parameters, thus drawing a sharp separation line between zero and non-zero coefficients. These methods are however strict theoretical devices that we cannot strongly rely upon.

A Lasso estimation is performed for the outcome (Granger Caused) variable on all the control variables, and of the treatment (Granger Causing) variable on all the controls as well. Then a post-selection least squares estimation of the outcome is performed on the treatment variables and all the control variables selected in at least one of the two Lasso estimations. Note that the controls correspond to all the other individuals (banks in our context) on which we are not performing the test for Granger Causality. Clearly the ratio is to capture all the relevant control variables, for each one of the treatment and outcome separately, and to only rule out those coefficients that are not relevant in either two of the selections.

The use of *Belloni* post double selection method proves to substantially diminish the afore mentioned omitted variable bias, otherwise present in the post-single selection, still being able to correctly detect the sparsity pattern and allowing for interpretability on the treatment variable when facing a context of High dimensionality. It also ensures with almost certainty the errors of the model obtained after OLS are orthogonal w.r.t the treatment variable (exogeneity).

In a VAR context the post double selection method can be formalized as follows:

Let $x_{GC,j}$ with j = 1,..., N_x where N_x = pN_j , denote the j-th column of X_{GC} (p is the number of lags), and consider the partial regressions:

a) $Y_I = X_{-GC}\ \gamma_0 + e_0$
b) $x_{GC,j} = X_{-GC}\ \gamma_j + e_j$

Note that in this context Y_I represents the outcome variable, X_{GC} is the treatment variable (or block), these are the variables we want to investigate Granger Causality about. X_{-GC} represents all the available control variables that we are not directly testing in GC relationship (all the others).

The two partial regressions above still represent a High Dimensional setting in our VAR context. We'll perform Lasso estimation method on both a) and b) in order to collect all the relevant controls, before performing the final least squares estimation of Y_I on X_{GC} and on all the relevant variables selected in the previous double selection procedure.

The omitted variable bias w.r.t the coefficients of X_{GC} would be originated where it exists a variable in X_{-GC} that has a non-zero coefficient in both regressions a) and one of the regressions in b). If the coefficient of such a x_{-GC} variable is zero in regression a) it means that it does not affect the outcome and hence not wrongfully omitted. If the coefficient is zero in all the regressions in b) it means that this control is not related to any variable in the treatment and omitting, it wouldn't result in a bias. Meaning the Lasso procedure would have to wrongfully rule out a non-zero parameter in both regressions a) and b) in order to create omitted variable bias (the probability of this happening gets asymptotically negligible). *(Margaritella et al.)*

The innovation is to perform variable selection useful for inference (selecting the controls to be included in the GC testing, for some treatment variable), without incurring in the bias generated by just using one Lasso selection passage.

The procedure in easy steps:

-First select a set of control variables which are useful to predict the treatment of interest (regress treatment "X" on all available controls using LASSO), helps identifying controls strongly related to the treatment thus helping to identify possible confounding.

-Then select additional control variables, choosing the variables that predict "Y" (the dependent we're modelling in our main model, VAR), this helps avoid omitted variable bias.

(These two steps will allow us to gather all the relevant variables, minimizing omitted variable bias) Given that we essentially have N_{GC} = NJ × NI × p steps of selection, it would be more appropriate to refer to this method as "post-NGC-selection".

-Estimate the coefficients of interest (of the treatment X on the dependent Y) by linear regression of Y on treatment X and on all variables selected in the first two steps. This in our context corresponds to OLS estimate the VAR (p) model on the variables (time-series) retained by post double selection.

Networks and Communities (Data and results):

Data

The main characteristic defining a stationary time series process, is that the variable tends to revert always to the mean, and the process is invariant when moved arbitrarily over time, so the properties of the series do not depend on the time at which the series is analysed. Such that no time trends can be identified, and modelling a subsection of the data corresponds in some way to model the whole-time horizon (appears like the series when located above its mean is subject to some "pressure" to go down again and cross the mean, to then be subject to a upward pressure when located below it). Non-stationary processes on the other hand are subject to upwards or downwards trends that are characteristic of the period considered, meaning that modelling one period instead of another makes a difference. Is our duty, to correctly estimate the model, to understand what the long-time behaviours of the variable are. The first problem arises when actually identifying whether a process is stationary or not; in fact, since we're dealing with variables evolving over time, how can we be really sure that a variable showing a strong upward trend in its history, will not in fact revert to its mean when evolving over the next time periods? It is impossible for us to certainly forecast the future, but with some degree of confidence we can say that some processes show non-stationarity because of the theoretical background that data come from. For example it is reasonable to expect an AR(1) process on GDP of one country will show non-stationarity, and will in fact tend to grow over time without inverting the trend. The main issue however, when trying to model nonstationary data, is that we have a high risk of detecting spurious correlations. The trend nature of a particular variable, when regressed on one or more variables that follow the same time trend, will cause our estimates to always generate high significance level and prediction capacity, just because the model is able to identify the same trend pattern followed by data, but without really being able to detect the real "co-movements of variables". Instead, in the case of stationary data the behaviour of the variables can be taken "in a vacuum" and estimates can be carried on with less risk of finding spurious relations. Of course the issue arises in cases such the one in our analysis, when dealing with high dimensionality and ultimately with a lot of variable that we don't know ad priori whether subject to a causal relation or not (since is our aim to find out whether this relation exists or not); so, ruling out causal relationship wrongly detected by the model, by means of a conceptual approach (like knowing that decreases in car crashes are not causing increases in GDP) is not a feasible approach when dealing with HD VAR models. We must be careful when identifying the time dependence of data and the way time, as sort of a control variable in this case, is affecting all the other variables in our problem. That's why most of the times we're just interested in modelling stationary time series data, or we make our non-stationary data stationary by performing manipulations as taking first differences. The exception to trying to always model stationary data is when we find ourselves facing variables that are "co-integrated".

Checking for cointegrated time series processes allows to spot nonspurious correlations: two series with integration of order 1 (I(1) meaning they're nonstationary but their first differences are), can be cointegrated (a linear combination of them leads to a trend with integration level zero), only if there exists a real relationship between the two. We can define the difference between the two time-series Yt-B*Xt, if a B exists such that this process is stationary, we can call B the cointegration parameter (or example if B is 1, means Yt-Xt is stationary, but Xt and Yt are alone nonstationary). Cointegration can be viewed as the mean reverting property of the difference between these two variables, such that the more the two variables are apart the more the mean reversion "tendency" is strong, like a rubber band pulling them back together whenever they start to deviate from each other; while two not cointegrated time series will not show this tendency to avoid distancing from one another. The important property is: cointegrating parameter can be estimated consistently via OLS by regressing one variable on the other (Xt on Yt as normal OLS, the marginal effect will in this case be our cointegration parameter), without having the problem of facing wrongly identified spurious relations ONLY if the two variables are cointegrated (two integrated I(1) series which are not directly causal related may none the less show significant correlation, spurious as explained above, and which is the main problem of trying to model non stationarity). It is clear how in general understanding the nature of our time series data is of great importance before proceeding with any estimation; and testing for unit-root and cointegration is pivotal.

The validity of tests employed in detecting degree of integration or cointegration depend however on the correct specification of the DGP.

The model we're using can be employed on non-stationary data without need for cointegration test beforehand.

Even if we tested for the selection of the corrected length p=2, in order to avoid issues of spurious relationships we impose $p > d$, where d is the suspected maximum order of integration of our time series, (d=2 handles potentially nonstationary time series). This means when selecting the lag, we will always need to impose a number of lags that is at least equal or greater to the suspected degree of integration. Artificially augmenting the number of lags of the Granger causing variable over the suspected degree of integration, allows to avoid spurious relationships being highlighted without the need to test for any cointegration or unit root before, this is explained among others in Toda and Yamamoto 1995. We need to reach a balance between the lag length employed and the suspected degree of integration so that when running post double selection by Lasso, the model has the sufficient variables to perform first difference estimation. The R package provides us with a function that selects for our dataset the correct number of lags to be employed using Bayesian Information criteria. In our case the selected lag is 2, and we set our maximum degree of integration to 2 (reasonable to expect economic variables to be integrated at most of degree 2). Hence spurious causality shouldn't be a problem. The data we're modelling in the main results are 'levels' (not differenced into stationary), however in the end of the thesis we also run models on differenced data that are made stationary artificially to compare the results obtained.

In cases the time series length and of covariates have the same magnitude, the post double selection criteria tend to select too many variables (better explained at page 22). Thus, the function we run to perform post double selection also presents a lower bound parameter, that regulates the number of coefficients selected, allowing the model to stop selecting significant variables in the double lasso sparsity seeking, only after a certain share of the parameters was selected (in the models presented here after this lower bound is set at 50 %).

The dataset is made up of 88 variables (Globally Systemically Important Banks and other banks showing global systemic importance from around the world) observations correspond to the price of the stock in different denominations, daily, over weekdays from 04/01/2005 to 17/03/2021 (for a total of 4227 observations). Variables are expressed in different units of measure, namely different currencies; however, this should not affect our analysis in finding patterns of causal relationship, as these can be detected from the movements more than the nominal values.

We obtain a network of 87 x 87 relations of spillover effect/GC relation, for 88 individuals.

<u>Results</u>

Our aim is to identify the existence of communities of individuals, and their evolution over time. Generally, we're interested in understanding the shape of the bank network (the extent and location of GC relationships among banks' stock prices) and how it changes in time. More precisely: the way the network and the communities changed in different economic-financial moments of the last 17 years. Thus 14 models employing post double selection method are run for each year from 2005 to 2021. The world, and most of others the western economies, was subject to different crises and subsequent reactions by central banks that brought to various economical moments, ranging from post 2008-2011 austerity measures and post covid-19 QE and other measures of expansion. Different periods of prosperity or recession, and the corresponding measures adopted by central banks, brought private banks to commonly enjoy the gains from Quantitative Easing for example, or, on the other end suffer the losses from the withdrawal of investments or measures of liquidity containment. The network of GC allows us to detect the extent to which the banking industry is tight in shared movements and reaction to changes in the economy (through the price of stocks over time, which however is a good depiction of this dynamics). As such, we can capture whether the system is so tangled up and "interdependent" to just expect shocks affecting one bank to have widespread effects on the whole network, or whether communities sharing isolated causal relations exist: such that shocks affecting a bank in a community will strongly determine the effects on other banks in that community. An arrow drawn in the network represents GC relation detected in the PDS model.

Decreasing the lower boundary on the selection process, or doing the same with the significance level, the models estimated over more years together does not increase in interpretability.

Figure 1 refers to the model estimated over 2019-21 simultaneously, Figure 2 refers to the model over 2005-07.

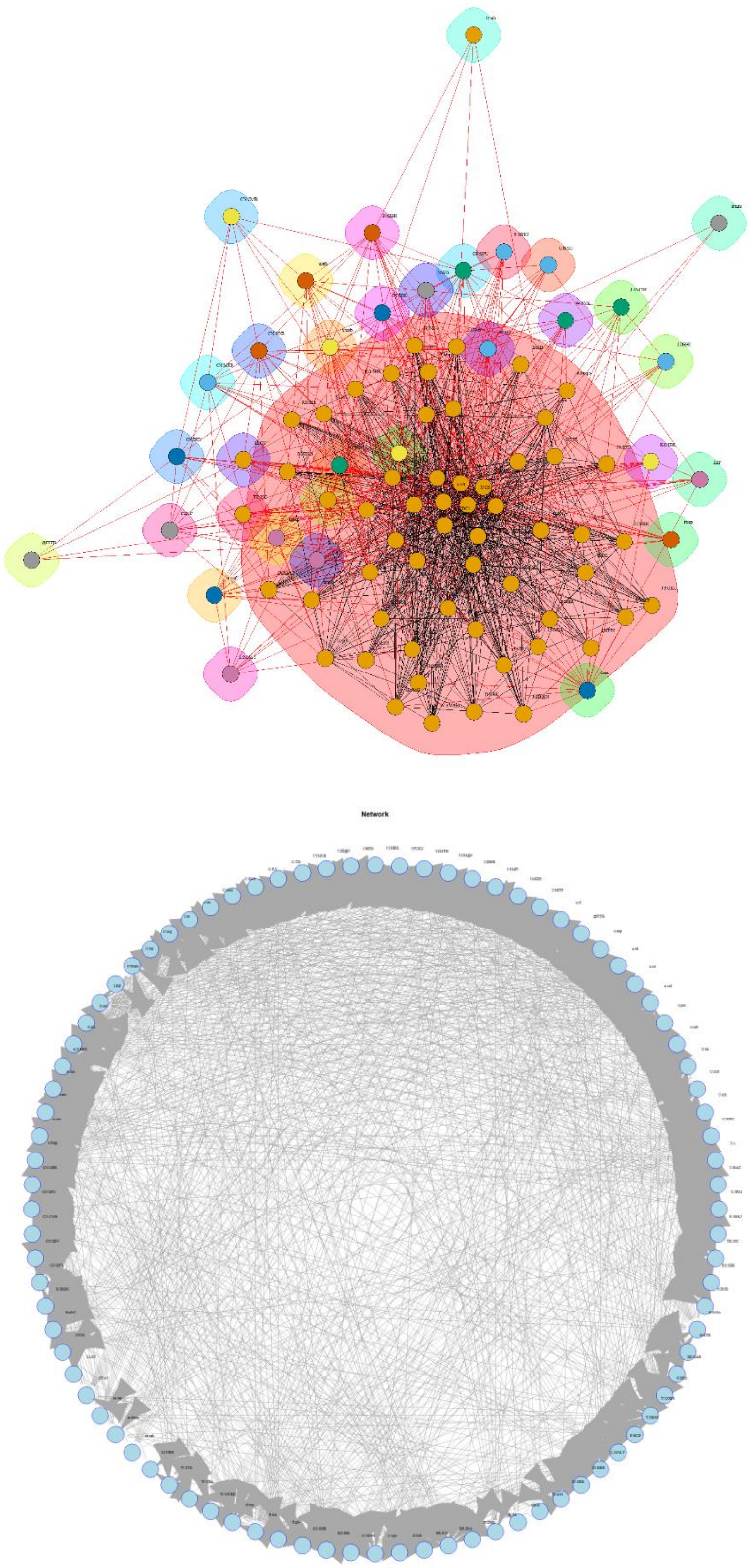
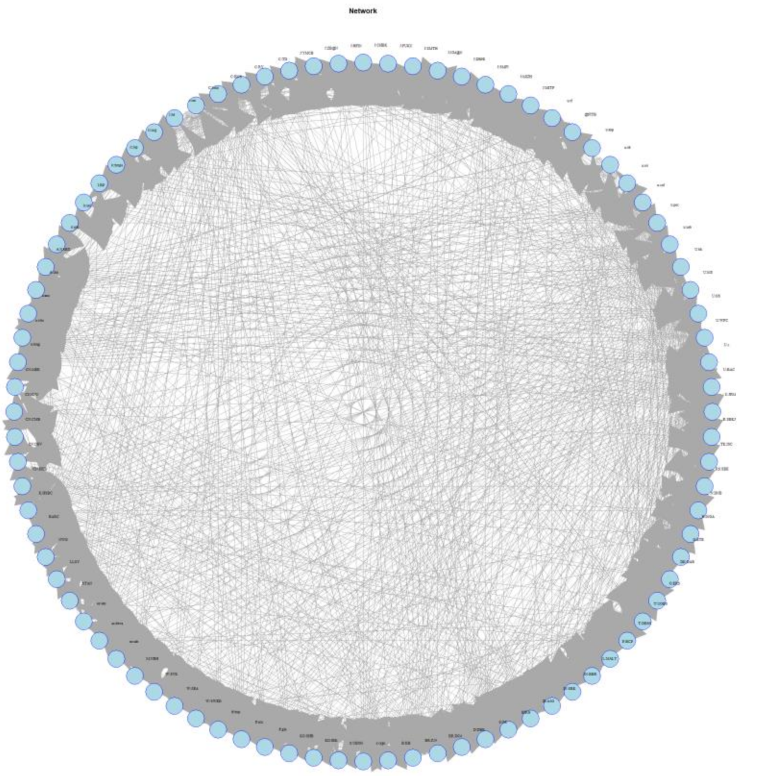


Figure 1



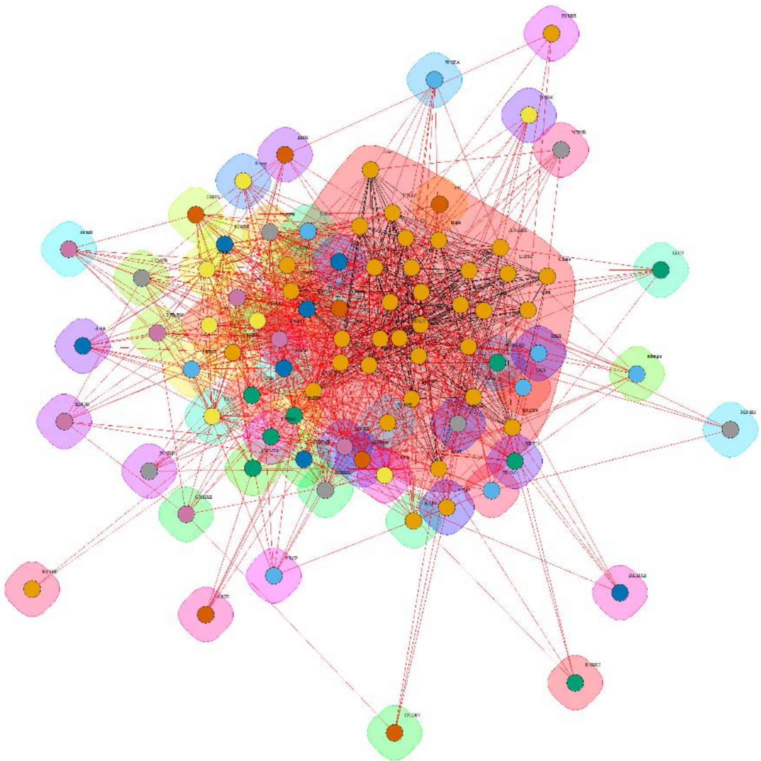


Figure 2

The networks show arrows of GC (predictability) occurring among and between almost all the individuals, and the communities being tangled, appearing as a big main community with just few individuals living on the border and showing fewer relationships. This may prove when considering time intervals of more years together the banking industry is deeply interconnected in a network of spillover effects.

The fact big commercial, investment banks, and particularly publicly traded banks (such as the ones belonging to our data pool), are subject to stringent regulations that, both previously and even more following the 2008 crisis, were put in place to decrease the amount of risk undertaken by financial institutions. Most of these big banks share a large pool of investment activities that they put their depositor's money into. These investments, however, are so differentiated that in aggregate terms they just tend to reflect in their outcomes the well-being of the financial system and of the economy. Even though the investments of banks are under constant inspection by the authorities to precisely avoid situations of systemic risk; this may cause the activities of all the big banks in the world to converge in the same direction to some extent especially during tumultuous times, explaining to some extent the behaviour we observe in the networks.

The situation however is different when it comes to smaller banks which are not part of our dataset, banks that are not publicly listed for example. These smaller financial institutions most of the times are inserted and deeply dependent on a specific economic/geographic area that their clients/depositors pertain to and do business in. Also, they're particularly active and focused on specific economic activities as serving a particular industrial sector in one country. This kind of banks are expected to enjoy a much bigger degree of "independence" and detachment from the rest of the banking industry. Analysing these smaller banks, we would rightly expect to observe more communities to show up and fewer causal relationships to be highlighted in the network, as each bank will be subject to different settings and environments that however will be much more stringent in determining the price of their stock.

Furthermore, publicly listed stocks are subject to movements of retail traders and other dynamics that will be likely to affect the whole banking industry through the stock prices, while for small banks the price of their stock will likely be determined mainly by the decisions undertake by the bank and the economic activities the bank is subject to. In our case, identifying a big main community, without being able to detect an evolution of smaller communities of bank over time, may just be a confirm of this conceptual intuition about the nature of the banking industry.

Still, the most reasonable cause for observing such dense networks over different years would be the effects of having a lot of temporal observations, or a lot of observations in general. Two observations should be pointed out: on one hand one may be led to think the number of observations when increasing will lead to a higher amount of information (which seems always a good thing for the researcher).

A trade-off exists between the amount of information we can fit into a model and the inference that can be done upon that model. In a sense, under Granger conception of causality for example, which is described in *Eichler 2013*, the causality relationship is always defined conditionally on an information set, which Granger defines as all the information available in the universe; since collecting all the information in the universe is an impossible task, the notion could be readapted to including all the possible treatment and control variables that we have at disposition, thus the more variables are included the better the chance of really capturing true causal relationships among individuals (from which our need to perform post double selection and use sparsity imposing methods). The situation however changes when dealing with the number of (temporal) observations: when performing the last step in our model, namely identifying significant coefficients of VAR model for GC in an OLS step, the significance of variables is defined by an F test. The F statistic is defined by a critical value which increases with the difference “n-k” so that when “n” is much bigger than “k”, the test will likely produce critical values always falling in the rejection area and thus always lighting up a significant coefficient. The result is that when we try to fit models over a temporal horizon that exceeds the year, the effect described above takes place producing a highly interconnected and tangled network. Furthermore the model won’t be able to identify communities, as each individual is highly dependent on one another; almost all the coefficients seem to show significance. Conceptually this can be confronted with a usual linear regression in the case the number of observations equal the number of variables, hence a perfect fit can always be found; the principle is almost identical. This means the huge amount of informations, and in particular the number of observations (since by conceptual framework we are to involve as many variables as possible), leads the model to always be able to find significant GC relationships to occur among individuals.

The reason why we observe these tangled and confused results over multiple years, may also be that the model detects GC relationships when these occur even in a small portion of time and not persistently over the entire time window. Even if a GC relation between two individuals exist in only one of the years considered by the model, the model is going to highlight it, in the end always producing a dense network that never shows sparse relations. For this reason, we try to plot the evolution of the banks pool by running 17 different models, one for each year, and order them chronologically.

We must also pay particular attention to the correct interpretation of our model. We're dealing with a model for VAR analysis of GC; a coefficient being highlighted or not, thus the GC arrow being drawn or not, corresponds to detecting the predictability of current values (stock price) of a particular individual conditional on the values assumed by the Granger Causing individual in the previous period (and on all the other "relevant" individuals). In this case having two lags on the previous two days over the whole year, (or more than one depending on the number of lags selected, which in our case is 2 as selected by BIC). What the network is telling us is the degree of predictability conditional on previous observations.

The density of the network, which can be easily understood from a picture, represents the degree of interconnection of the industry, or more precisely the degree of predictability of individual banks conditionally on the others. Observing the evolution of the network over the years allows to associate the changes in the density and the shape with the economic and financial condition respectively of each year. The networks and communities depict a particular pattern of response to financial events (such as great recession or covid pandemic). One could use the networks and its density as an indicator of financial stress and use it to detect whether an industry is forecasting a financial crisis. Furthermore, the degree of density and the existence of communities itself is an indicator of financial well being of an industry, and it can be used to forecast spillover effects of the failure of a specific bank for example by using concepts offered by "Graph theory".

Graph theory offers different instruments helpful in identifying useful measures to study our spillover networks. For example, we can define a centrality measure that allows to identify the relative importance of nodes and edges. Eigenvector centrality identifies nodes that tend to be most frequently reached. Still there exist different ways to analyse centrality measures depending on the characteristics of the network we're dealing with and on the kind of analysis we want to perform. We may be interested in studying the dynamics of the network or the robustness of the network when a link is removed (dynamic importance of node removal is of great interest and importance). The concept of non-conserved spread is used to study changes to a complex network when, contrary to the conserved case, the source of "excitement" in the network is possibly infinite; this is used and best suited to study situations such as the transmission of infectious diseases, neural excitation and spread of information and rumour. This framework of non-conserved spread could be also employed in our analysis to study the effect of a financial crisis hitting on the network and the response (determine a measure of Stress withstanding of the banking industry). In the same way it may be useful to identify so called "hubs", nodes that are subject to a larger number of links.

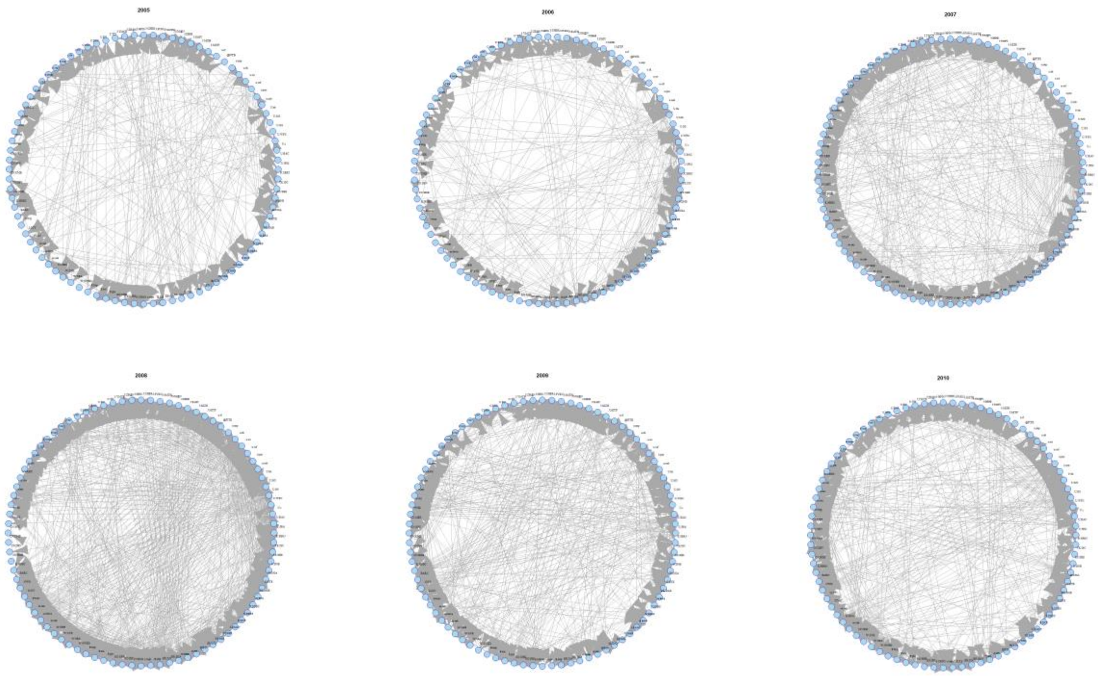
We could use some of the above methods to answer: "what are the institutions more endangered?", "which are the subjects that may cause systemic risk to arise in the industry and cascade effects on the whole banking sector?". In this report of results, we're just interested in depicting a general interpretation of the networks evolution and of the pool of banks in its integrity.

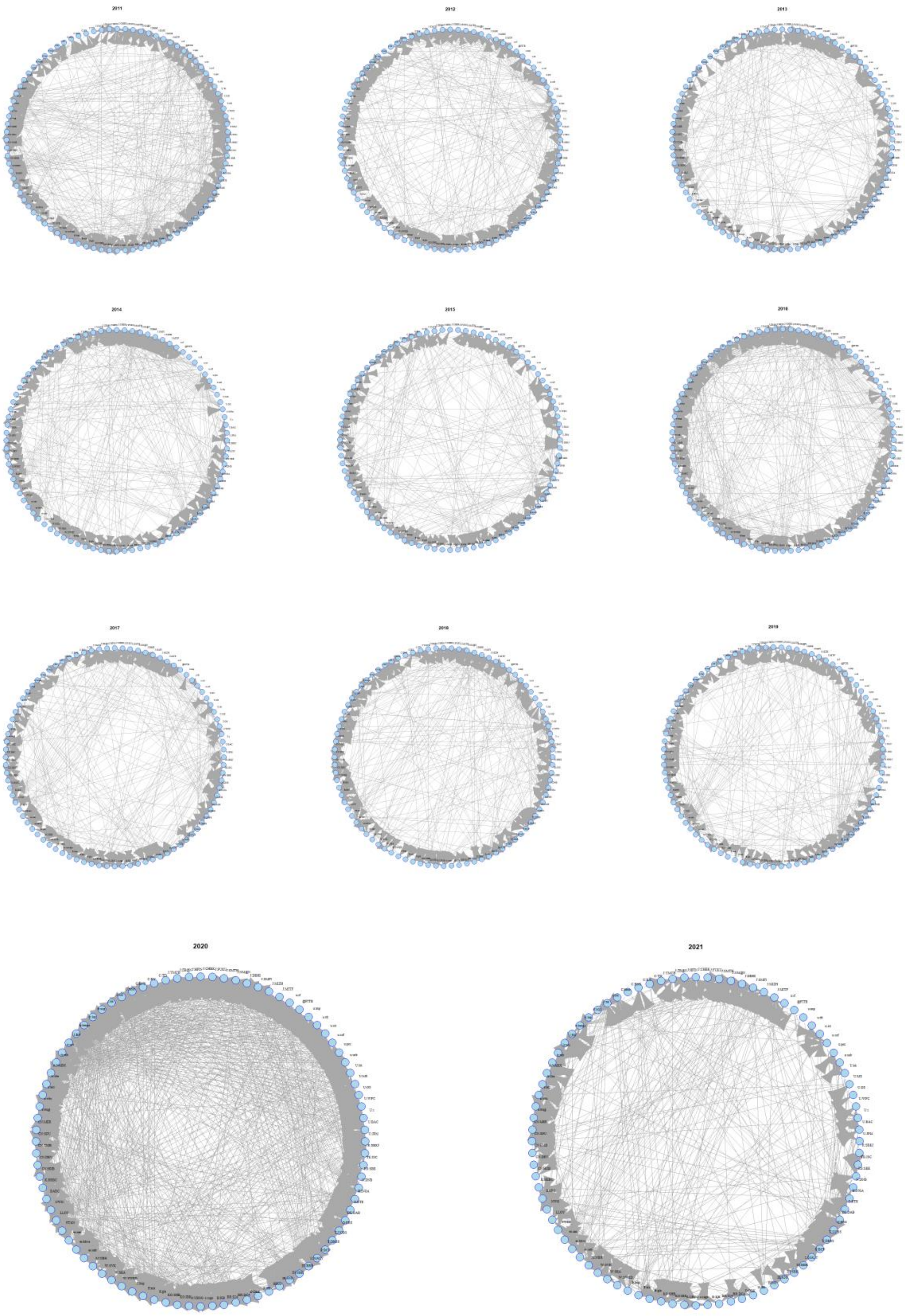
By a visual analysis of the plotted networks evolution over the years we can understand how the level of interconnection and predictability within the network varies over time. Then we can then compare the behaviour assumed by the network with the economic and financial events of the period. The idea is to identify a pattern that the network assumes in financial crises and that can be detected to forecast in advance the degree of systemic risk or the envisage of such crises by the banking industry.

It appears that the communities are much more shaped, we can identify different ones, in years that are fair from financial crises, namely years fair from 2008/09 Great Recession or the 2011/13 European sovereign debt crisis and 2019/20 covid-19 pandemic. The communities melt into just one big group with some outliers when in the years struck by crisis. We can clearly see the same evolution over time when observing the networks, with the number of arrows, defining GC relationships detected by the model, increasing significantly when moving towards the crisis, and diminishing when moving away. The network assumes an extremely tangled and hardly interpretable shape in the years of the interested by financial turmoil.

This could be interpreted (as suggested before) as the crisis hitting all the industry in the same way and forcing it to abide the same strategies and employ the same solutions to thrive. And as the degree of financial interconnection and co movements increasing making it much more likely for other banks to suffer losses from other banks. To provide an example we can think of bank runs being much more likely for all other banks when one bank fails, as the public trust in banks collectively diminishes with a fast-spreading fear bank failure. In this way the behaviour of one bank will likely affect much more the behaviours of one or more banks, in what we may define a “tension” arising. In the same way when the industry is calm and thriving the banks adopt different strategies and sectors of interest, developing ties among one another which are likely to shape into communities, each one bearing interdependence relations that they share (for example we can sometimes observe communities made of banks that belong to the same country). Networks are showing less dense as the causal relationships become fewer and better defined. By the definition of Granger Causality, we can observe over time how the degree of predictability of one individual bank stock price increases conditional on the lagged stock prices values of all the other individual banks in the chosen sample.

Below is presented the evolution of networks and communities in the years 2005 to 2021. Figures 3 and 4 show the evolution of communities in the same years with a gap (2011-2017).





The idea behind plotting communities is that: we're able to detect certain groups of elements which share more connections internally than with the rest of the network, this helps in a visual interpretation of the evolution over time.

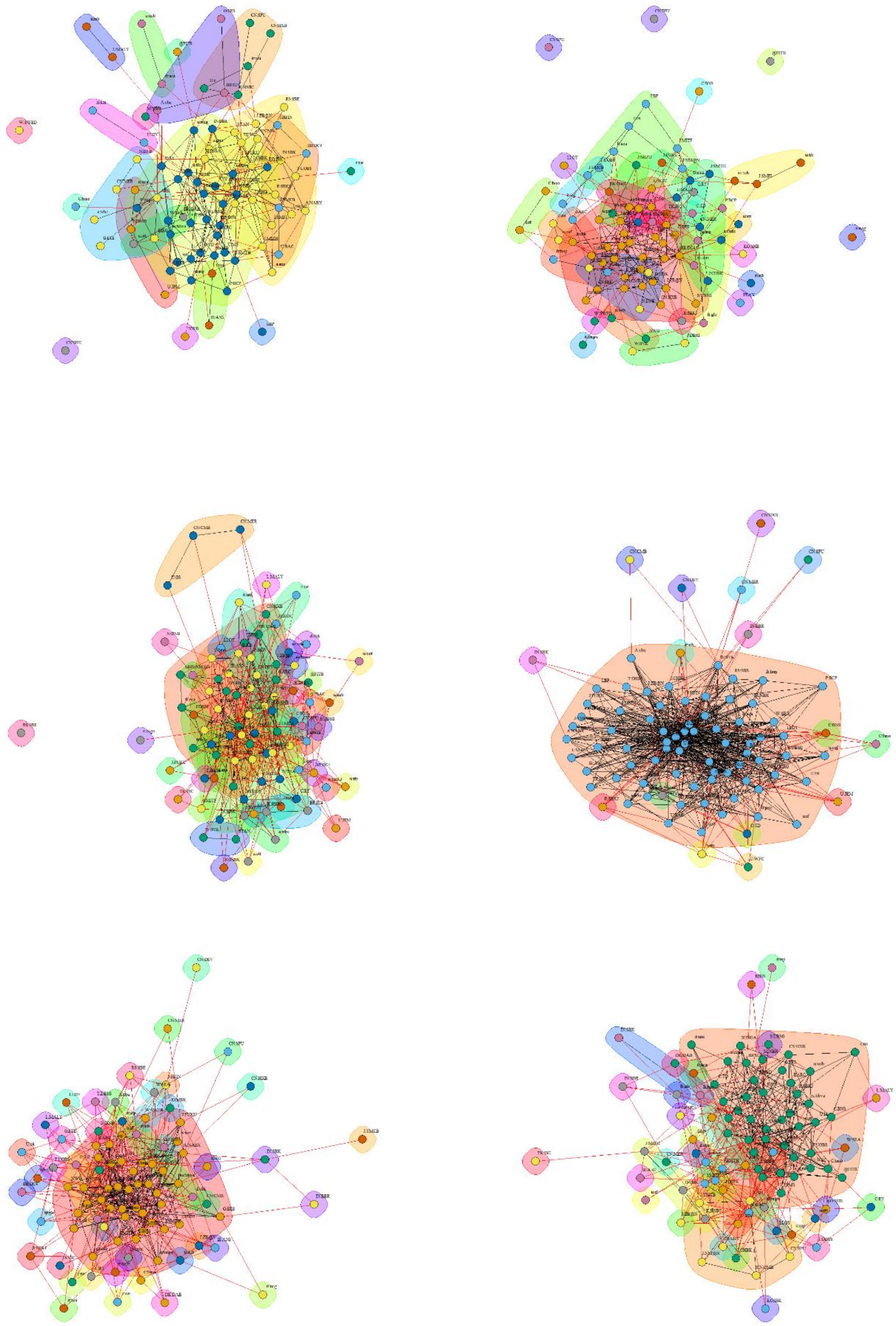


Figure 3 (2005 to 2010)

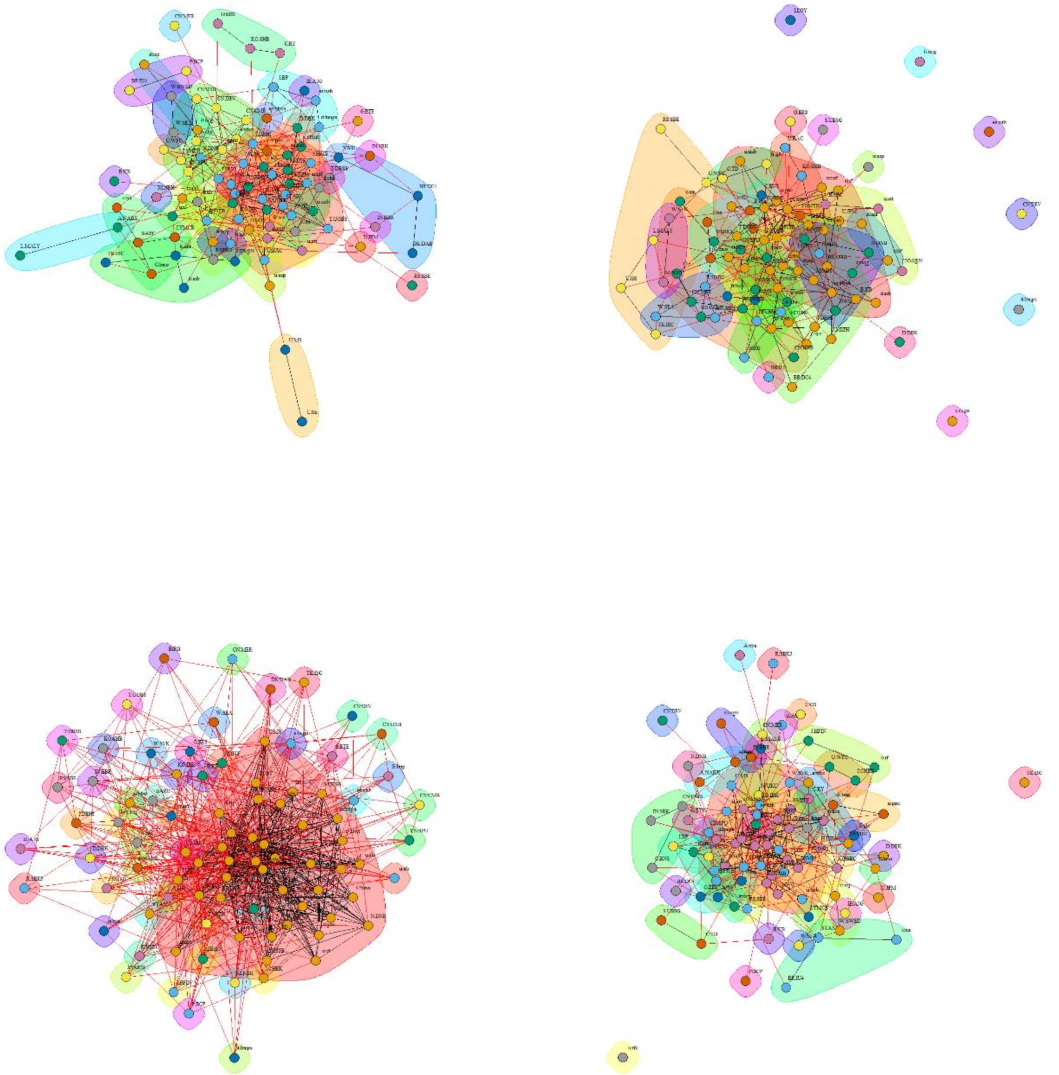


Figure 4 (2018 to 2021)

Conclusions:

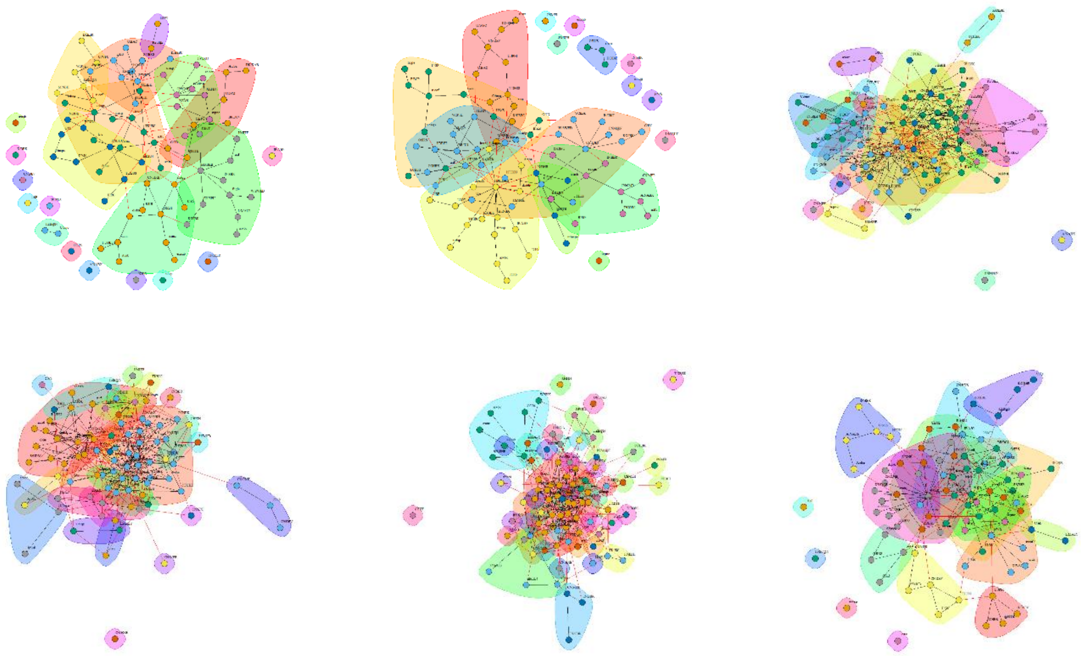
The use of alternative estimation and variable selection methods to impose sparsity, namely the Post Double Selection method proposed by Belloni et al. and discussed by Margaritella et al. , allows the detection of Granger Causality relationships in High Dimensions (after the due theoretical premises) with a sufficient degree of certainty about the goodness and solidity of results obtained, and avoiding the omitted variable bias entailed by the use of post-single selection. In a context of Vector autoregression analysis over a pool of 88 banks this method allowed to shape the evolution of a spillover network over 17 years, with the objective of understanding the evolution of the network and obtaining a marker/indicator of financial crises to be detected from the network itself. Overlapping the results with the economic and financial macro events of the time periods we can observe the tendency of the spillover effects to become stronger under the crisis showing an higher level of entanglement and interconnection; this can be interpreted as the degree of predictability of the whole sample (banking industry) to increase in the crisis (conditional on “itself”) , probably because of the same macroscopic events that the industry is subject to require the same reaction by all the banks and are luckily to affect all of them in the same way (and most of the cases in the same direction). It’s important to stress how the nature of this results rests strongly on the nature and nominal size of these banks, which make them subject to some common elements (most importantly the degree of control and the economic activities that they’re subject to). Also consider that even a formal detection of Granger Causality is not enough to define a true causal relation, and we may in fact think of it as predictability. The main advantage of the post double selection framework is to consistently handle high dimensionality, so that even with few temporal observations with respect to variables we’re able to estimate a model. Taking smaller time periods of some months or some weeks can allow to build a forecasting measure: an index of industry interdependence could be built on the varying number of connections detected in the network of spillovers. So that, resting on our results, an alert for financial crises can be triggered by the increasing degree of interdependence and by the disappearance of isolated communities. A suggestion for further research would be to perform the same analysis dealt with in this short paper on a pool of regional banks; subject to some theoretical/conceptual premises to be discussed by the researcher. This may give the chance to detect more precisely and verify the presence of Granger Causal relations, also being able to contextualize the pool of banks and their main activities (sometimes extremely differentiated from one regional bank to the other) and to understand the way a situation of distress affected that specific economic or geographic area that the dataset is referring to, and how systemic risk was originated.

References:

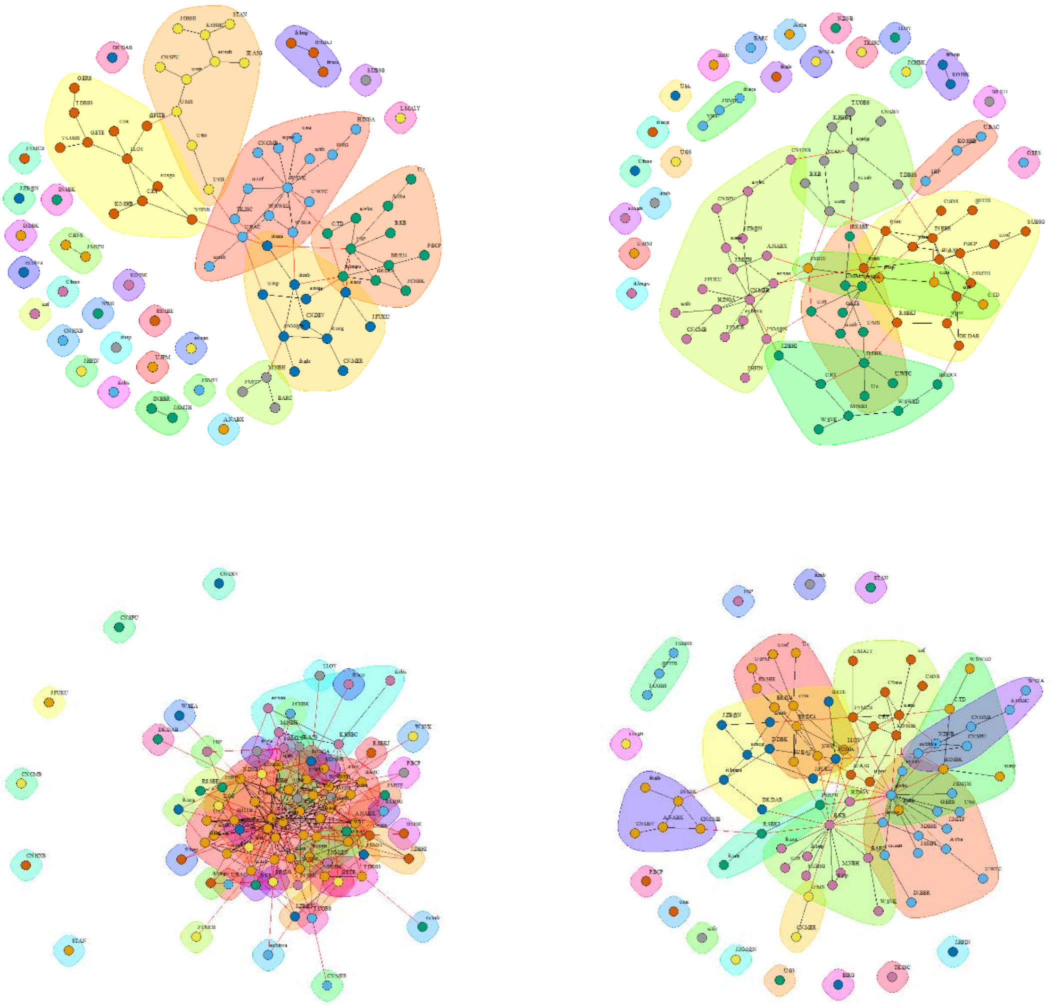
- Baek, C. and Park, M. (2021) 'Sparse vector heterogeneous autoregressive modeling for realized volatility', *Journal of the Korean Statistical Society*, 50(2), pp. 495–510. doi:10.1007/s42952-020-00090-5.
- BELLONI, A., CHERNOZHUKOV, V. and HANSEN, C. (2014) 'Inference on Treatment Effects after Selection among High-Dimensional Controls', *The Review of Economic Studies*, 81(2 (287)), pp. 608–650.
- Chernozhukov, V. *et al.* (2022) 'High-dimensional Data Bootstrap'. Available at: https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsarx&AN=edsarx.2205.09691&site=eds-live&scope=site (Accessed: 9 May 2023).
- CORSI, F. (2009) 'A Simple Approximate Long-Memory Model of Realized Volatility', *Journal of Financial Econometrics*, 7(2), pp. 174–196. doi:10.1093/jjfinec/nbp001.
- Donohue, J.J. and Levitt, S.D. (2001) 'The Impact of Legalized Abortion on Crime', *The Quarterly Journal of Economics*, 116(2), pp. 379–420. Available at: https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=edsjsr&AN=edsjsr.2696468&site=eds-live&scope=site (Accessed: 9 May 2023)
- Eichler, M. (2013) 'Causal inference with multiple time series: principles and problems', *Philosophical Transactions: Mathematical, Physical and Engineering Sciences*, 371(1997), pp. 1–17. doi:10.1098/rsta.2011.0613.
- Hecq, A., Margaritella, L. and Smeekes, S. (2021) 'Granger Causality Testing in High-Dimensional VARs: A Post-Double-Selection Procedure', *Journal of Financial Econometrics* [Preprint]. doi:10.1093/jjfinec/nbab023
- Thierry Tressel *et al.* (2022) *The Global Bank Stress Test. [Elektronisk resurs]*. International Monetary Fund. Available at: https://search.ebscohost.com/login.aspx?direct=true&AuthType=ip,uid&db=cat07147a&AN=lub.7201795&site=eds-live&scope=site (Accessed: 9 May 2023)
- Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society - Series B 58, 267–288.
- Toda, H. Y ad T. Yamamoto (1995). Statistical inferences in vector autoregressions with possibly integrated processes. Journal of Econometrics, 66, 225-250.
- Kock, A. B. and L. Callot (2015). Oracle inequalities for high dimensional vector autoregressions. Journal of Econometrics 186 (2), 325–344.

Stationary data

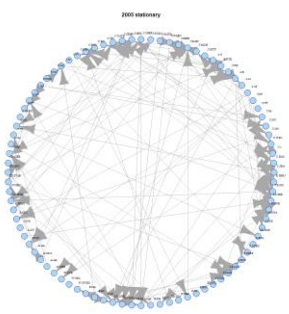
We run the same models that we reported in the main results, but now instead of running levels (raw variables), we difference the data into stationary. We can observe the networks show overall the same behaviour over years as the networks computed using nonstationary data. Namely showing the spillovers increasing and the communities losing shape to reach a peak under the crisis of 2008 and 2020, and to then decrease and reshaping in well-defined communities when moving from the crisis. We can see the model better identifies communities, shaping them in a much more precise way, and the number of edges in the network (of GC relations/ spillover effects selected) is significantly smaller than those identified over nonstationary data. Networks are much less tangled and more interpretable. The lower number of edges and significant GC relations is what makes the identification of the communities much easier and better defined, as it becomes easy to observe nodes sharing many connections together and few connections with the rest of the network. The model can group together in the same communities banks sharing the same country of origin, which is a great result. For instance, we can observe in 2005 the Japanese banks (J:) all being closely connected one another, and Chinese banks (CN:) showing the same behaviour in years from 2007 to 2010, differently from those of other nations.

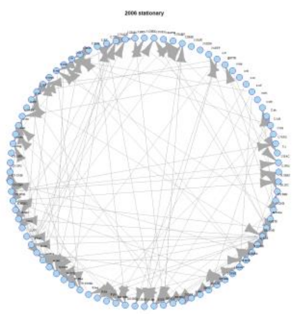


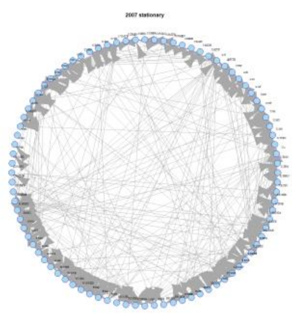
Above are Communities 2005-2010 from up left corner moving right

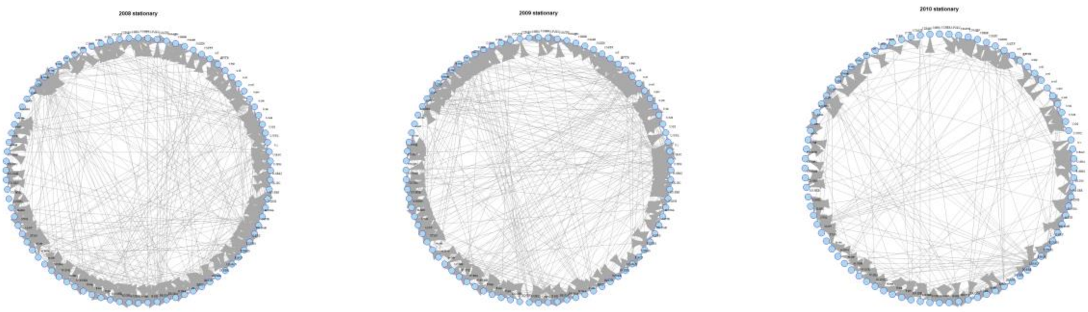


Above are communities 2018-2021

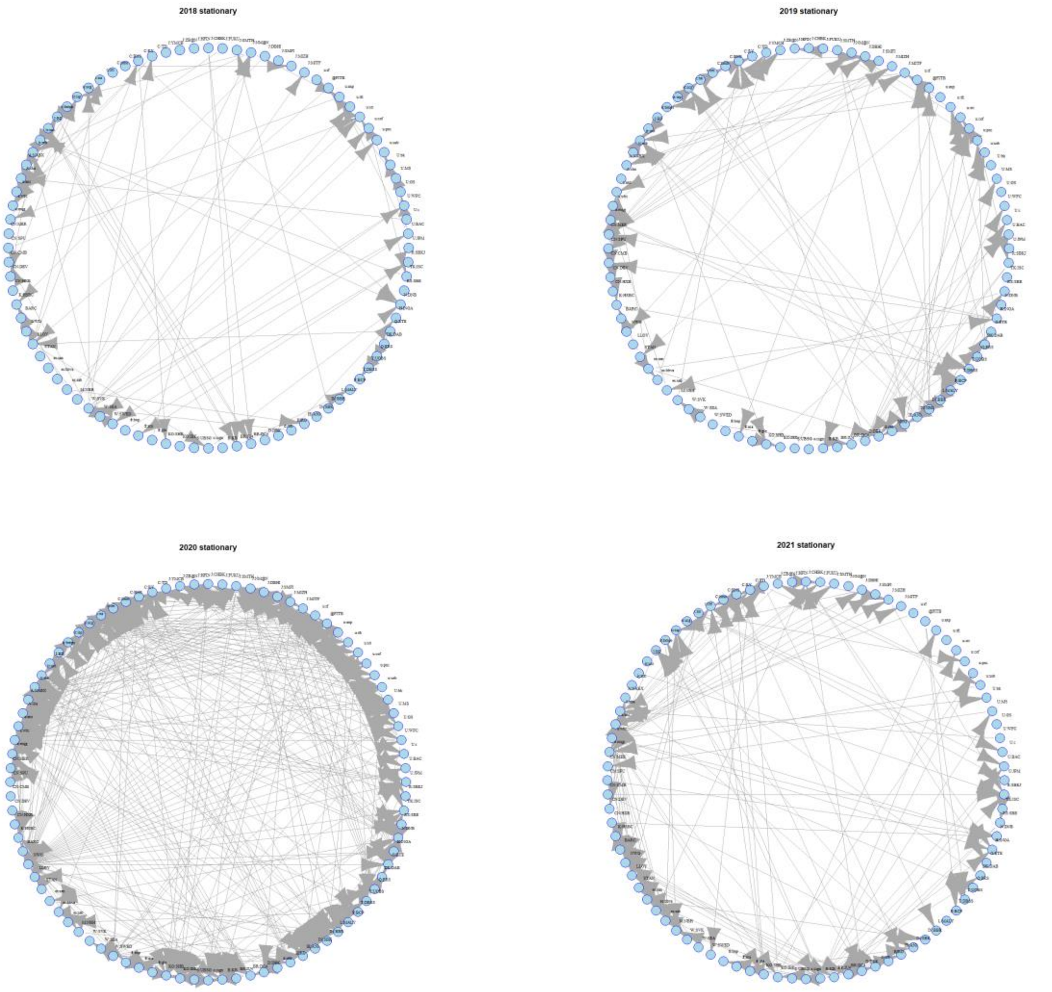








Above are Networks 2005-2010



Above are networks 2018-2021