



Combining permutation tests to rank systemically important banks

Lorenzo Frattarolo¹ · Francesca Parpinel¹ · Claudio Pizzi¹

Accepted: 12 October 2019 / Published online: 28 October 2019
© Springer-Verlag GmbH Germany, part of Springer Nature 2019

Abstract

In this work we propose the use of a nonparametric procedure to investigate the relationship between the Regulator’s Global Systemically Important Banks (G-SIBs) classification and the equity-based systemic risk measures. The proposed procedure combines several permutation tests to investigate the equality of the multivariate distribution of two groups and assumes only the hypothesis of exchangeability of variables. In our novel approach, the weights used in the combination of tests are obtained using the Particle Swarm Optimization heuristic and quantify the informativeness about the selection. Finally, the p value of the combined test measures the reliability of the result. Empirical results about the selection of G-SIBs show how considering the systematic (β), stress ($\Delta CoVaR$) and connectedness components (*in–out connection*) of systemic risk cover more than 70% of weight in all the considered years.

Keywords Systemic risk · Global Systemically Important Bank · Particle Swarm Optimization · Permutation tests

1 Introduction

The framework of Systemically Important Financial Institutions (SIFIs) was introduced by the Financial Stability Board (FSB) in October of 2010; these institutions were defined as those “whose disorderly failure, because of their size, complexity and systemic interconnectedness would cause significant disruption to the wider financial

✉ Claudio Pizzi
pizzic@unive.it

Lorenzo Frattarolo
lorenzo.frattarolo@unive.it

Francesca Parpinel
parpinel@unive.it

¹ Department of Economics, Ca’ Foscari University of Venice, San Giobbe Cannaregio 873, 30121 Venice, Italy

system and economic activity” FSB (2010). The methodology to select which Globally Systemically Important Banks (G-SIBs) should have additional capital requirements according to the Basel III agreement is outlined by the Basel Committee on Banking Supervision (BCBS), BIS (2013). The Regulator builds the selection process on annual data disclosed by banks of member countries. The collection of the dataset is complex and time-consuming, furthermore the selection is usually published one year later. The use of slow varying balance sheet annual data, on the one hand, increases the stability of the ranking, but on the other hand, it cannot promptly detect sudden changes. Such a delay could result in sub-optimal monitoring of institutions and inefficient policies to mitigate the systemic risk.

In this respect, several works have proposed systemic risk measures based on market data; for a comprehensive review see Silva et al. (2017) and Benoit et al. (2017), which show what scholars and practitioners use to rank the systemic importance of each institution in a timely manner. However, according to Silva et al. (2017), many published articles are focused on specific measures and consider different samples of institutions. Therefore, a fruitful comparison of those measures is impossible. Remember that Silva et al. (2017) said: “It would be interesting if there were more comprehensive and comparative studies of measures of systemic financial risk, applying them to the same set of banks, discussing the advantage and disadvantage of each approach, and checking where they clash and where they complement each other; this could enhance the monitoring of financial institutions in an interesting manner”. This comparison was partly performed in Giglio et al. (2016) to investigate the macroeconomic meaning of global systemic indicators. To our knowledge, only the paper by Benoit et al. (2017) performs the proposed comparison on the same sample at single institution level, but the focus here is on the systematic nature of some of the considered measures.

In general, it is still an open question if these measures are really relevant in the classification performed by Regulators. This ranking is important not only for the additional capital requirements imposed on G-SIBs but also for consequences on their market value as outlined by Moenninghoff et al. (2015).

We answer this question by statistical means using a combination of randomized permutations tests and proposing a new weighting scheme for the combination based on Particle Swarm Optimization. Our completely nonparametric choice stems from nature of systemic risk measures. In fact being based on collective behavior or tails of distributions their unknown cross sectional distribution forbids the use as explanatory variables for common parametric binary choice models. The proposed methodology requires only the exchangeability hypothesis of single systemic risk measures. Exchangeability is a reasonable assumption for each random variable considered in the cross section of institutions, which, even in the presence of some heterogeneity, could be considered distributed as mixtures of i.i.d. and thus exchangeable random variables. In addition, using a randomized permutation test we have no small sample issues, even with the limited number of institutions we consider.

Of the most used measures of systemic risk according to Silva et al. (2017), we restrict our analysis to the ones based on stock market returns, which is the most easily accessible kind of financial data. In our analysis we use the following nine:

- β , which is the linear regression coefficient of the bank excess return r_i over a market equity index excess return r_M , and is used to disentangle documented systematic effects in the other measures according to Benoit et al. (2017).
- the *Marginal Expected Shortfall* (MES) introduced by Acharya et al. (2010) which is the average return of a bank during the worst 5% of days for the market; MES is able to predict a bank's contribution to a crisis.
- The $\Delta CoVaR$ proposed by Adrian and Brunnermeier (2016), which is the difference between two Conditional Values at Risk (CoVaR), that is the market conditional quantile with respect to the bank return quantile. The difference is between the CoVaRs computed conditioning on the bank return being equal to its median and its 5% quantile; the conditional quantiles are obtained through quantile regression.
- Five network centrality indices, that are *in connections*, *out connections*, *in + out connections*, *1/closeness* and *eigenvector centrality*; these network centrality measures may also be applied to a Granger causality network, according to Billio et al. (2012).
- *PCA* proposed by Billio et al. (2012), which is obtained by considering the principal components of over all the examined banks and is related to the absorption ratio of Kritzman et al. (2011) and Kinlaw et al. (2012).

To visualize the multivariate nature of the problem and understand how our method can deal with it, let us consider the marginal Empirical Cumulative Distribution Functions (ECDFs) in the non-G-SIBs and G-SIBs groups; the previously mentioned considered measures are computed from 2015 data. Figure 1 shows the ECDFs for each risk measure in the two groups established by the FSB. It is clear that for some variables, such as β , *MES* the main differences between the two groups arise in the central part of the distribution; in other cases, the two groups differ more at low values, i.e., *in-out connections*, or at high values, for instance $\Delta CoVaR$ and *PCA*. Thus, we argue that combining the information gathered from different systemic risk measures makes our test about the equality of the two FSB groups more accurate.

In this work we propose a methodology based on a combination of nonparametric permutation tests that consider several measures of systemic risk to reproduce the ranking of the Regulator and allows us to investigate the relationship between those timely academic measures and the annual Regulators indicator.

Understanding this relationship has the outcome of substantiating the relevance of equity-based measures for systemic importance and to put into light the potentiality of our proposal to allow timely preliminary monitoring of those institution individuated by the relevant measures. In particular, our methodology allows us to detect the relative weights of equity-based measures in explaining the Regulator designation. In our application we will consider the same assessment sample of banks used by Regulators. In doing this we also contribute a comparison of the most-used equity-based systemic risk measures, in line with the suggestion of Silva et al. (2017). Given the large number of measures proposed in the literature, we are far from exhaustive in our selection of the specific measures. Our choice is based on a trade-off among popularity, uniformity of the data used and simplicity. Even if our selection of measures is subjective, our

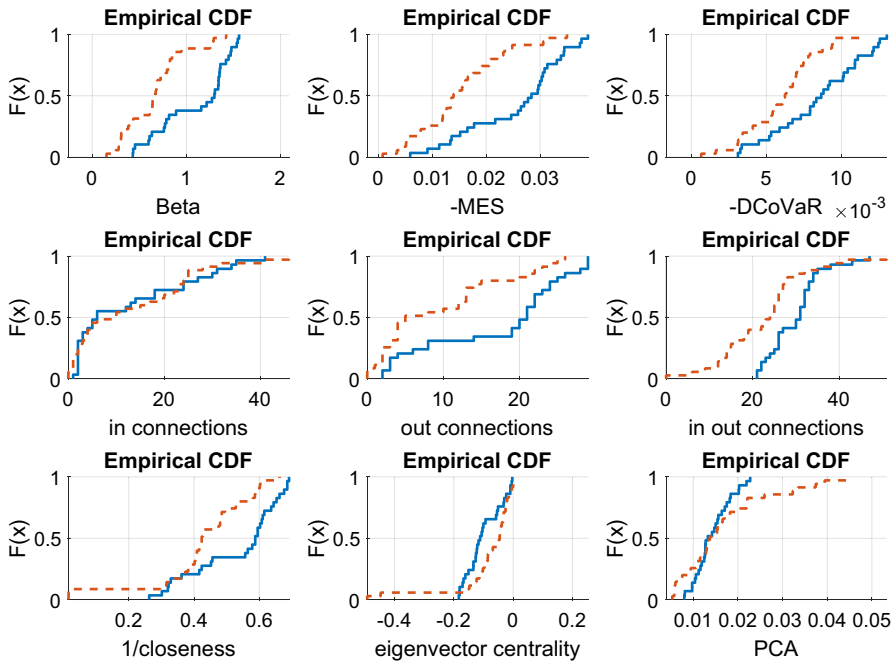


Fig. 1 ECDFs of the involved risk measure; year = 2015. Solid blue line: G-SIBs, Dotted red line: no-G-SIBs (colour figure online)

methodology is not, and a nonparametric multivariate test can, in principle, be applied jointly to any number of measures.

The multivariate nature of our procedure is important because each measure intercepts only some of the features of the complex phenomenon known as systemic risk, as we have previously pointed out. This complexity is implicitly recognized by the Regulator, which uses 12 indicators in its G-SIBs designation as shown in Table 1 and then discussed in Sect. 2.

A well-known way to solve our inferential problem is to combine statistics for several variables and use permutation tests, see Winkler et al. (2016) and references therein. In this technique, each variable is weighted. Rejection of the null hypothesis of equality between groups implies that our results are adherent with the Regulator choice.

The novelty of our approach is in choosing the combination weights by minimizing the p value of the combination test in such a way that the weights add up to one. In this manner, the vector of weights, or the solution of the constrained optimization problem, can be interpreted as the relevance measure for the FSB choice. In addition, we can interpret the optimal p value as a measure of reliability for the procedure. Additionally, we reconstruct a fictitious G-SIBs selection based on each single macro category of indicators used by the Regulator, using a simple heuristic. In this way, we are able to repeat the procedure and to compute the relevance weight of academic measures relative to *size*, *interconnectedness*, *substitutability*, *complexity* and *cross-*

Table 1 Indicators and relative score weights used by the Basel Committee for the evaluation of systemically important banks

Category	Indicator	Indicator weight (%)
Size	Total exposure	1/5 = 20
Interconnectedness	Intra-financial system assets	1/15 = 6.66
	Intra-financial system liabilities	1/15 = 6.66
	Securities outstanding	1/15 = 6.66
Substitutability/financial institution institution infrastructure	Payment activity	1/15 = 6.66
	Assets under custody	1/15 = 6.66
	Underwritten transactions in debt and equity markets	1/15 = 6.66
Complexity	National amount of OTC derivatives	1/15 = 6.66
	Trading and AFS securities	1/15 = 6.66
	Level 3 assets	1/15 = 6.66
Cross-jurisdictional activity	Cross-jurisdictional claims	1/10 = 10
	Cross-jurisdictional liabilities	1/10 = 10

jurisdictional activity. Non-smoothness of the p value as a function of weights requires a global optimization method. The parallelizable Particle Swarm Optimization (PSO) is a valid candidate, see Kennedy and Eberhart (1995). We overcome the computational complexity by exploiting the multi-core architecture of the System for Scientific Computing of Ca' Foscari (SCSCF).

The paper closer to our approach is Li and Tseng (2011) where the same minimization of p value is proposed but the value of the weights are restricted to be binary and an exhaustive search is used. This restrict the application to less than 10 variables.

The paper is organized as follows. In Sect. 2 the methodology is presented, in Sect. 4 we show the results for the FSB selection and for macro-categories-driven selections, and in Sect. 5 we discuss the results and propose some possible extensions.

2 Methodology

The institutions included in the analysis for each year are the ones in the BCBS assessment sample. The requirements for inclusion in the assessment sample are the following:

- Banks that the BCBS identifies as the 75 largest global banks at the end of the financial year, according to the leverage ratio exposure measure based on Basel III.
- Banks that were designated as G-SIBs in the previous year (unless supervisors agree that there are compelling reasons to exclude them).
- Banks with a score produced by the indicator-based measurement approach that exceeds a cut-off level set by the committee.

- Banks that have been added to the sample by national supervisors using supervisory judgement (subject to certain criteria).

The banks included in the assessment sample must disclose the informations needed to compute a composite score that considers different aspects of systemic risk: *size*, *inter-connectedness*, *substitutability*, *complexity* and *cross-jurisdictional activity*. Those macro categories are themselves composites and a finer subdivision comprises the measurement of the 12 variables, detailed in Table 1 BIS (2014). The score is a weighted sum of the indicators with the weights designated in the last column of Table 1 representing the score.

We optimize each selection proposed with combination of permutation tests Pesarin and Salmaso (2010). Let X be a cross-sectional variable observed on x_i with $i = 1, \dots, n$, and n is the sample size. The statistical units, corresponding in our case to different banks, are divided in two groups, g_1 and g_2 , according to a given selection criterion. Now we can compute the two-sample Kolmogorov–Smirnov statistic i.e., $KS(\hat{F}_1, \hat{F}_2) = \sup_x |\hat{F}_1(x) - \hat{F}_2(x)|$. We denote the observed value of this statistic with v_{obs} . The choice of a Kolmogorov–Smirnov statistic could be replaced by an alternative equivalent distributional statistic such as the Cramer–Von-Mises or the Anderson–Darling Statistics. According to the permutation test principle Pesarin and Salmaso (2010), we compute an approximated p value using

$$P_B = \frac{1}{B} \sum_{b=1}^B I(v_b \geq v_{obs}) \quad (1)$$

where $I(\cdot)$ is the indicator function and equal to 1 if the condition in parenthesis is satisfied and zero otherwise, and $v_b = KS(\hat{F}_{g_1^b}, \hat{F}_{g_2^b})$ is the value of the statistic calculated on the b th permutation. In a multivariate framework that considers p variables we may verify the null hypothesis of equal distribution using a permutation test for each one, and we call this a partial test. We denote each observed partial statistic by $v_{j,obs}$. In order to test the null hypothesis of global equal distribution in the two groups, we have to aggregate the p values of all partial tests in a single test using a combination function. In our case, we use the Fisher *omnibus* function with weights w_j :

$$t = - \sum_{j=1}^p w_j \log(\lambda_j). \quad (2)$$

where λ_j is the p value of the j th partial test.

Therefore, according to Pesarin and Salmaso (2010), given B random permutations, we can obtain an approximated p value for the combination by

$$P_B(w) = \frac{1}{B} \sum_{b=1}^B I(t_b \geq t) \quad (3)$$

and denoting α as the significance level of the global test, if $P_B(w) \leq \alpha$, we reject the global null hypothesis of equal distribution of the two groups at α significance level. Note that the value of $P_B(w)$ depends on the set of weights used in the Eq. (2). In addition, the value of $P_B(w)$ gives information about the intensity of the difference between the multivariate distribution of the two groups; indeed, the higher the significance of the test, the deeper the difference. In order to obtain a good global index able to catch the differences, we choose to optimize the weights w_j , $j = 1, \dots, p$, minimizing the p value subject with the constraint that the weights add up to one.

In our procedure, the aim to minimize the p value is reached by optimizing the Fisher omnibus function, which combines several tests where the discretionary elements are the weights w_j , $j = 1, \dots, p$. The complexity of the problem suggests to use a numerical global optimization method. In particular we consider the bio-inspired iterative meta-heuristic called Particle Swarm Optimization (PSO) introduced by Kennedy and Eberhart (1995).

One of the great advantages of the PSO is that it is highly parallelizable. In fact, the updating of particles' characteristics is independently computable given the previous iteration. However, one of the main drawbacks of the applied PSO is that it is stochastic in nature. In our case, also the permutation test is also randomized, doubling the source of stochasticity. We performed several analyses to evaluate the robustness of the whole procedure. The results obtained show negligible differences with several setups reaching a precision of p values below the fourth decimal digit, see Frattarolo et al. (2016). Robustness checks reported in the aforementioned paper ensure reproducibility of the whole procedure. The use of 200 particles and 10,000 random permutations suffices to obtain reliable results.

As pointed out by an anonymous referee and confirmed by simulation studies available upon request, the procedure shows some sensibility to collinearity or strong association among the explanatory variables. For this reason before the application of the methodology we pre-select explanatory variables according to the following scheme. We computed 3 measures of association, i.e. Pearson's correlation index, Spearman's correlation index and Kendall's tau. We prioritize keeping simpler measures for each type, according to the list: β , *in + out connections*, *PCA*, *in connections*, *out connections*, *-MES*, *-\Delta CoVaR*, *eigenvector centrality* and *1/closeness*.

We start deleting variables with at least one of the association measures higher than 0.65 with β , from the remaining we remove the ones with at least one of the association measures higher than 0.65 with *in + out connections*, and so on so forth till we eliminate all the variables with all association measures higher than 0.65 among themselves.

The final methodological step is designing an out of sample procedure to understand the effectiveness of our optimally weighted nonparametric test. We choose to benchmark the model against the equally weighted test and the most used parametric binary choice models, i.e. logit and probit regressions. A comparison with Li and Tseng (2011) approach is left for future research.

We estimated the parametric models, in sample, using weakly associated variables and then selecting the regressors till everything is significant at the 0.1 significance level. We then performed an out of sample exercise using the optimal weights, logit and probit parameters estimated with the data of the previous year before to understand

the discriminatory power on the choice of the current year. For our procedure and the usual combination with equal weights we use as a measure of discriminatory power the p value of the test. For logit and probit we instead compute the predicted probability for the two classes of the current year and use the p value of a permutation test based on Kolmogorov Smirnov two sample statistic. The KS test is consistent with our methodology, but it is also a common synthetic measure of discriminatory power being the minimum distance from the diagonal of the Receiver Operating Curve Campbell (1994).

3 Dataset description

As we explained above, we are interested in two different analyses. The former is devoted to study how a restricted number of systemic risk indices are able to reproduce the ranking of the Regulator and how the weights of the combination change over time, the latter investigates the relevance of weights of the same indices relative to each single macro category of the indicators used by the Regulator. Since the disclosure of macro categories was not mandatory in 2012, and since we were not able to retrieve the macro-category indicators for all the assessment sample banks from the balance sheet disclosed on the Bank of International Settlement website due to linguistic barriers, the datasets differ slightly in the two kinds of analysis.

The sample of G-SIBs comprises the 75 largest banks as determined by the Basel III leverage ratio exposure measure, BIS (2013). Each year, the names of the institutions in the assessment sample are disclosed on the BIS website. Information concerning the capital requirement buckets are extracted from documents published by the FSB each year announcing the G-SIBs. The stock returns of the assessment sample banks are obtained from Bloomberg. Note that as some banks are non-quoted companies, they are not included in our analysis; thus, we do not have the data corresponding to every institution in the assessment sample. In particular, the sizes of analyzed sets are 66 out of 75 banks in 2012 and 2013, 68 out of 75 banks in 2014 and 64 out of 75 banks in 2015. The market returns are represented by the MSCI World Index. In order to analyze the relevance of the risk measures in the Regulator decisions, we perform the optimization procedure for each year.

The disclosure of the macro-category indicators composing the systemic risk score consists of the publication of the balance sheet of the assessment sample banks on the BIS website from 2013 onward. For some institutions data are available in a more convenient format. In particular, for European banks, the measure is available from the European Banking Union (EBA) in *excel* format; the Federal Financial Institutions Examination Council's (FFIEC) provides these measurements in the same format for US banks. Data for the designed G-SIB is available in *excel* format by the Office of Financial Research. For non-G-SIBs non-EU or US Banks, we conducted an incomplete data collection.¹ The institutions for which the data are missing are the ones for which we were not unable to overcome linguistic barriers in reading the balance

¹ The analyzed dataset is available on request.

sheets. Accordingly, the final sample analyzed is composed of 41 banks out of 75 in 2013, 45 banks out of 75 in 2014 and 59 banks out of 75 in 2015.²

4 Empirical results

We followed the procedure outlined in Sect. 2 to pre-select non associated systemic risk measures. There are some regularities in the hystorical pattern of association that exclude some variables. First, we confirm how β and *Marginal Expected Shortfall* are almost perfect substitutes through the whole analysis. β is strongly associated with *PCA* in 2012 and with ΔCoVaR and *out connections* in 2014 and 2015. Moreover *1/closeness* and *eigenvector centrality* are also perfect substitutes, and they are also strongly associated with *out connections*. β and *In + out connections* are always included and *PCA* is excluded only in 2012 in line with the statistical interpretation given at the end of Sect. 1. In fact, we have one variable covering the centre of the distribution (β) one variable that covers the lower tail (*In + out connections*) and another fourth from a variable that covers the upper tail (*PCA*).

Figure 2 summarizes the results of the procedure illustrated in Sect. 2 and applied to the G-SIB designation. The graph depicts the weights adding up to 1 for each year from 2012 to 2015, coloured according to each corresponding risk measure. In addition, we report in parenthesis the optimal p value as a measure of analysis reliability. The importance of β is evident for all the period considered testifying the difficulty of disentangling systematic from systemic risk. *in-out connection* is relevant for all the years, and *eigenvector centrality* has also its role several times. These results suggest the inclusion of *Too Interconnected to Fail* interpretation of systemic risk as a part of the narrative. The ΔCoVaR has a strong role in the first year, then it overlaps with β and get excluded by our initial collinearity procedure. Since 2012 and 2013 saw the developing of the European Sovereign Debt Crisis we can interpret the ΔCoVaR as the unexpected stress to the system, that adds information during turmoils. Finally *PCA* seems to gain relevance whenever *eigenvector centrality* loses its traction, but does not seem. In general, the analysis suggests that considering the systematic (β), stress (ΔCoVaR) and connectedness component (*in-out connection*) of systemic risk cover more than 70% of weights in all the years.

Now we want to study how each risk measure is important to Regulator decision for each macro-category indicator.

To fulfil this aim, we consider the G-SIBs institutions selected by the Regulator for each category and compute the minimum value of the macro-category indicator. Then, we include all the institutions with values higher than the minimum in the new G-SIBs group based only on this category. The group analysis for each macro-category indicator is reported in Figures 3, 4, 5, 6 and 7. Please note that for each macro-category the obtained optimal p value is sufficiently low to validate the importance of the category in the Regulator's decision, except for *substitutability/financial institution infrastructure*.

² The complete list of included institutions is available upon request.

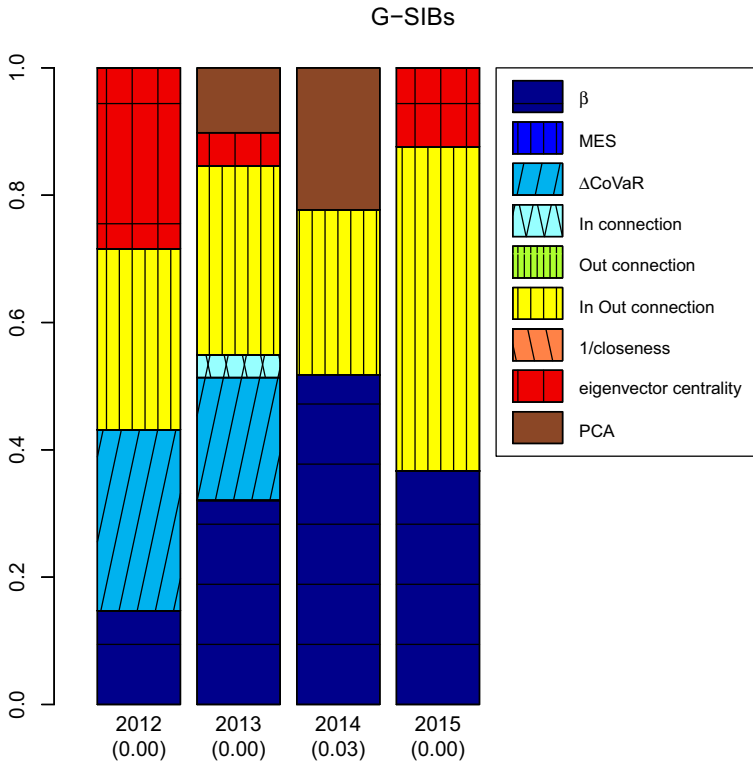


Fig. 2 G-SIBs selection. p values in parenthesis

The *size* measures the *total exposure* of a bank with respect to counter parties all over the world as used in the Basel III leverage ratio. Since there is high heterogeneity in exposure it is difficult to have a stable decomposition of this category in terms of our systemic risk variables. As shown Figure 3, we obtain significant p values only for the last year where all the weight is allocated to *eigenvector centrality*.

As detailed in Table 1 *Interconnectedness* is composed of three indicators: *intra-financial system assets*, *intra-financial system liabilities* and *securities outstanding*. The natural candidates for explaining this category are the interconnectedness measures, however, we must provide some remarks to understand this relationship.

First of all, we do not include other kinds of financial institutions in our systemic measure sample and therefore we include only a part of the complete banking system, even if it is designed as the most important. Second, it is worth noting that our measures are statistical in nature and include a broader set of influences among banks than just those caused by a transaction among them. Considering these facts, we indeed find a stable contribution from *in-out connection*; additionally, we have an important contribution of $\Delta CoVaR$ in 2013 and half of the weight explained by *PCA* in 2014. 2015 is almost in its entirety in the domain of connections but it is not statistically significant. We also note that a simple measure of connectedness that counts the number of relationships is much more relevant than centrality measures of institutions, which

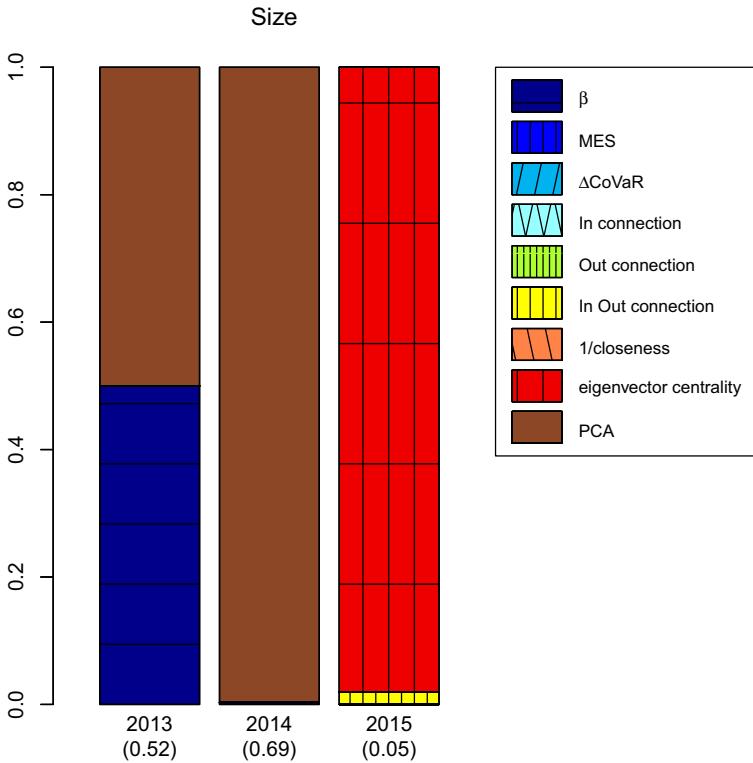


Fig. 3 The contribution of systemic measures to selections based on category size. *p* values in parenthesis

is in accordance with the fact that the indicators measure the number of transactions of a given institution and not its importance in the network of transactions.

Three indicators are used to measure *substitutability/financial institution infrastructure: assets under custody, payment activity and underwritten transaction in debt and equity markets*. The only result that we have here is the inadequacy of our set of systemic risk measure to capture this macro-category.

The *complexity* category is described using three indicators: *national amount over the counter, level 3 assets and trading and available for sale securities*. Connectedness measures are relevant also for this category, suggesting again that statistical connections measures are capturing different layers of connections. Additionally, again, ΔCoVaR takes the most weight in 2013, leaving space for *PCA* in the following years.

The last category *cross jurisdictional activity* is composed of two indicators: *cross-jurisdictional claims and cross-jurisdictional liabilities*. Here, we can see the same pattern as the previous category, but stronger, with *PCA* progressively replacing the ΔCoVaR contribution. A possible explanation of these findings is suggested by the global nature of the institutions in the sample, which cover most of the world. The institutions that play a major role internationally are also the ones that contribute the most to the first principal components. In this way, *PCA*'s weight increase could be read as a consequence of the increasing systemic importance of Chinese banks.

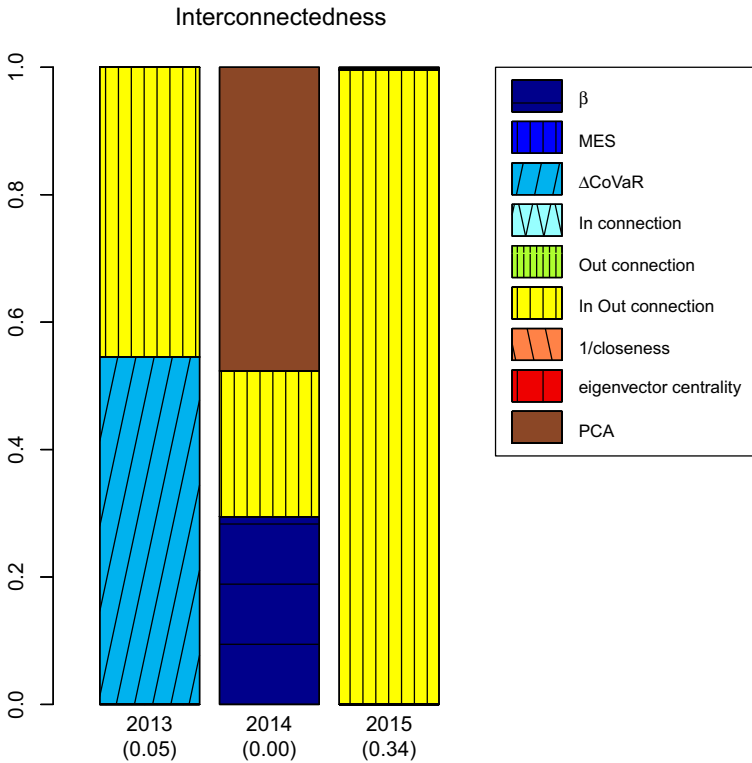


Fig. 4 The contribution of systemic measures to selections based on category *interconnectedness*. p values in parenthesis

4.1 Out of Sample

We performed the out of sample exercise as described in Sect. 2 for the G-SIBs selection and the macro-categories. We consider the alternative, equivalent to our proposal, if they differ by 1% or less. Results are robust with different choice of the 1% thresholds. For non equivalent outcomes the lower p value wins. We report in Table 2 the counts of each case.

Most of the times our model discriminatory power is equivalent to the one delivered by competitors. But parametric models are never better than ours, and we outperform 3 times over 4 the test with equal weights. Beating the parametric models is not surprising since given the variety of measures we use they are probably most of the times misspecified. Doing better than the test with equal weights, instead, is not foregone. Given the year by year variability that we showed in the macro-category in sample results, out of sample analysis could have preferred a more inclusive methodology as the test with equal weights.

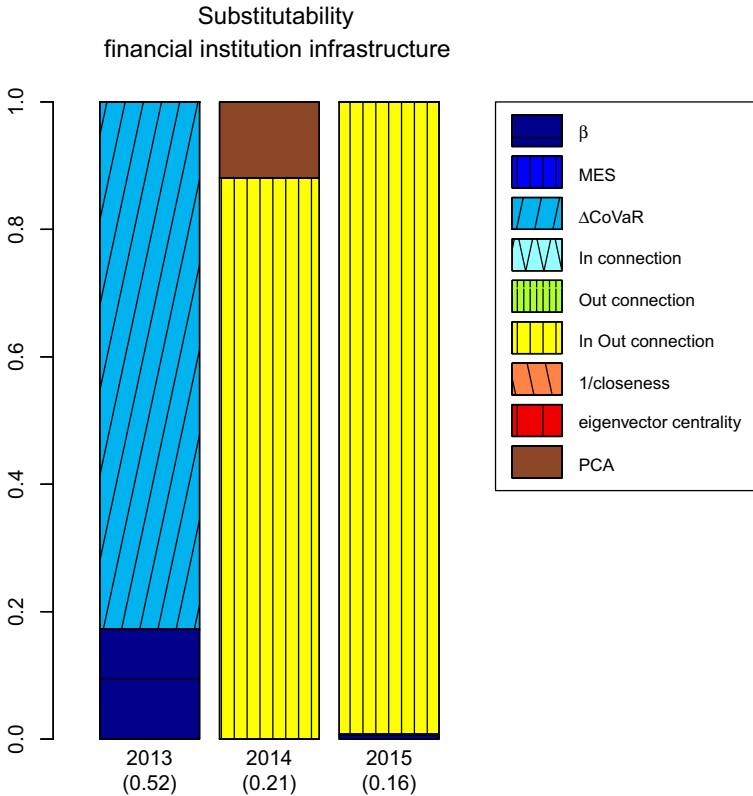


Fig. 5 The contribution of systemic measures to selections based on category *substitutability/financial institution infrastructure*. p values in parenthesis

5 Conclusions

The paper investigate the relationship between the Regulator's G-SIBs classification and the equity-based systemic risk measures using a non parametric combination of randomized permutation test, based only on the exchangeability hypothesis, with a novel weighting scheme based. The novel weights are obtained minimizing the p value with highly parallelizable PSO. They represent the relevance of the variable in the selection and optimal p value the goodness of the procedure. We have interesting results concerning the G-SIBs selection, as the systematic (β) stress ($\Delta CoVaR$) and connectedness components (*in-out connection*) of systemic risk cover more than 70% of weight in all the years considered. Our work then suggest that monitoring only those measures could provide a timely ranking of institutions. Nevertheless we are obliged to note a lack of regularity in the macro-category findings. Given the availability of data for only few years, it is difficult to understand where this variability comes from. Different categories are associated with different systemic measures in different years. Those results could be signal of a dynamic relationship among categories and systemic measures. In this regard, considering that 2012 and 2013 are deep inside the European

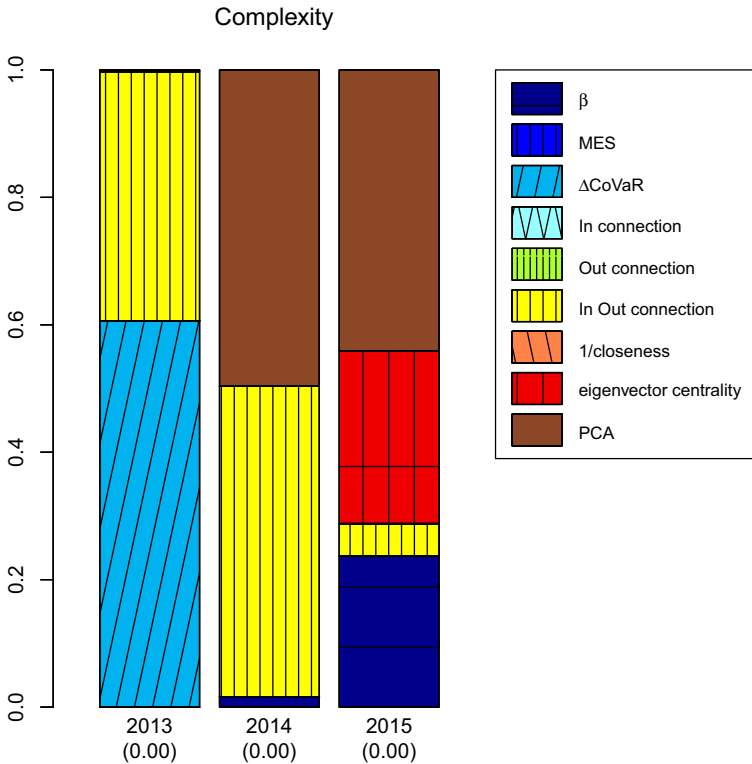


Fig. 6 The contribution of systemic measures to selections based on category *complexity*. *p* values in parenthesis

sovereign debt crisis, the important role of ΔCoVaR in only those year for the G-SIBs selection and the macro-categories could be a clear example of the dynamic behavior, by which some measures are informative only during turmoil. Even with this variability in the in sample results, we are able to out-perform in an out of sample exercise parametric models and an equal weight combination test.

We also outline the fact that equity-based systemic risk measures considered in this paper might be augmented and used to track changes in the decision variables used by the Regulator. In this light it would be important to also include different and more refined measures such as the measures proposed by Diebold and Yılmaz (2014) and recently extended by Demirer et al. (2015) to overcome the curse of dimensionality; these could now be computed with our sample. Analogous refinement of MES (Brownlees and Engle 2011; Acharya et al. 2012) and CoVaR (Girardi and Ergün 2013) based on multivariate GARCH were proposed. More broadly speaking, completely new equity-based measures are emerging from the literature and could be included in future. Additionally, a finer analysis based on the 12 indicators composing the macro categories, as suggested in the text, could, for example, shed additional light on the relationship among directional variables i.e., transactions vs. statistical connections. All those refinements and new measures are interesting venues for future research. This research would continue the quest to obtain more overlap of academia and Reg-

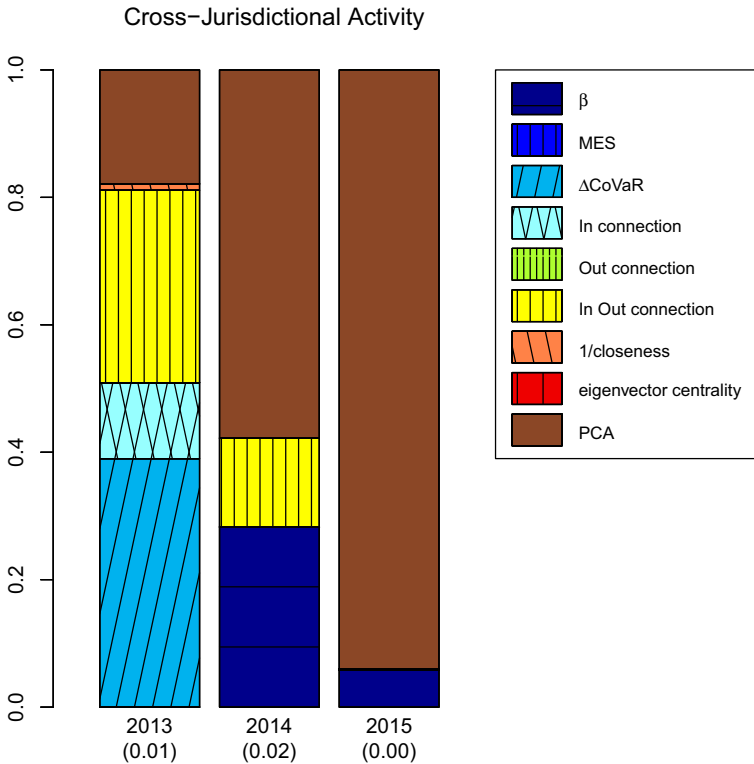


Fig. 7 The contribution of systemic measures to selections based on category *cross-jurisdictional activity*. *p* values in parenthesis

Table 2 Out of sample counts of proposed procedure with respect to equal weight, logit and probit, according to *p* values comparison

Ref. model	Equivalent	Underperform	Outperform
Equal weights	9	1	3
Logit	8	0	5
Probit	8	0	5

ulator assessment of systemic risk, and our proposed methodology could be a viable way of doing this. In addition, as a final remark we stress that the proposed original methodology is flexible enough to be applied in very different fields. For example, even narrowing our attention only to finance, it could be easily used to compare credit ratings with bonds yields or CDS premiums, or balance sheet ratios with stock performance measures.

Acknowledgements We wish to thank the Editor and two anonymous referees for very useful comments and suggestions which have helped to improve and develop the paper further. The authors are also grateful to The System for Scientific Computing of Ca’ Foscari (SCSCF) for computations.

References

- Acharya VV, Pedersen LH, Philippon T, Richardson MP (2010) Measuring systemic risk. Technical report, Department of Finance, NYU
- Acharya V, Engle R, Richardson M (2012) Capital shortfall: a new approach to ranking and regulating systemic risks. *Am Econ Rev* 102(3):59–64
- Adrian T, Brunnermeier MK (2016) Covar. *Am Econ Rev* 106(7):1705–41
- Benoit S, Colliard JE, Hurlin C, Pérignon C (2017) Where the risks lie: a survey on systemic risk. *Rev Finance* 21(1):109–152
- Billio M, Getmanski M, Lo A, Pelizzon L (2012) Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *J Financ Econ* 104:535–559
- BIS (2013) Global systemically important banks: updated assessment methodology and the higher loss absorbency requirement. <http://www.bis.org/publ/bcbs255.pdf>. Accessed 3 July 2013
- BIS (2014) The G-SIB assessment methodology—score calculation. <http://www.bis.org/bcbs/publ/d296.pdf>. Accessed 6 Nov 2014
- Brownlees CT, Engle R (2017) SRISK: a conditional capital shortfall measure of systemic risk. *Rev Financ Stud* 30:48–79
- Campbell G (1994) Advances in statistical methodology for the evaluation of diagnostic and laboratory tests. *Stat Med* 13(5–7):499–508
- Demirel M, Diebold FX, Liu L, Yilmaz K (2015) Estimating global bank network connectedness. Working paper 1512, Koç University–TUSIAD Economic Research Forum, SSRN 2631479
- Diebold FX, Yilmaz K (2014) On the network topology of variance decompositions: measuring the connectedness of financial firms. *J Econom* 182(1):119–134
- Frattarolo L, Parpinel F, Pizzi C (2016) Systemically important banks: a permutation test approach. *Rivista Italiana di Economia Demografia e Statistica* LXX:41–52
- FSB (2010) Reducing the moral hazard posed by systemically important financial institutions. http://www.fsb.org/wp-content/uploads/r_101111a.pdf. Accessed 11 Nov 2010
- Giglio S, Kelly B, Pruitt S (2016) Systemic risk and the macroeconomy: an empirical evaluation. *J Financ Econ* 119(3):457–471
- Girardi G, Ergün AT (2013) Systemic risk measurement: multivariate GARCH estimation of CoVaR. *J Bank Finance* 37(8):3169–3180
- Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the IEEE international conference on neural networks, vol 4, pp 1942–1948. <https://doi.org/10.1109/ICNN.1995.488968>
- Kinlaw W, Kritzman M, Turkington D (2012) Toward determining systemic importance. *J Portf Manag* 38(4):100–111
- Kritzman M, Li Y, Page S, Rigobon R (2011) Principal components as a measure of systemic risk. *J Portf Manag* 37(4):112–126
- Li J, Tseng GC (2011) An adaptively weighted statistic for detecting differential gene expression when combining multiple transcriptomic studies. *Ann Appl Stat* 5(2A):994–1019
- Moeninghoff SC, Ongena S, Wieandt A (2015) The perennial challenge to counter too-big-to-fail in banking: empirical evidence from the new international regulation dealing with global systemically important banks. *J Bank Finance* 61(Supplement C):221–236
- Pesarin F, Salmaso L (2010) Permutation tests for complex data: theory. Applications and Software. Wiley, West Sussex
- Silva W, Kimura H, Sobreiro VA (2017) An analysis of the literature on systemic financial risk: a survey. *J Financ Stab* 28:91–114
- Winkler AM, Webster MA, Brooks JC, Tracey I, Smith SM, Nichols TE (2016) Non-parametric combination and related permutation tests for neuroimaging. *Hum Brain Mapp* 37(4):1486–1511

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.