Journal of Financial Stability 72 (2024) 101221

Contents lists available at ScienceDirect

Journal of Financial Stability

journal homepage: www.elsevier.com/locate/jfstabil

Spillovers in Europe: The role of ESG

Karoline Bax [a,*], Giovanni Bonaccolto [c], Sandra Paterlini [b]

[a] Technical University of Munich, TUM School of Management, Campus Heilbronn, Center for Digital Transformation, Heilbronn, Germany
[b] Department of Economics and Management, University of Trento, Trento, Italy
[c] Department of Economics and Law, Kore University of Enna, Enna, Italy

ARTICLE INFO

Keywords:
ESG
Sustainability
Systemic risks
Network analysis
Financial stability

ABSTRACT

This paper explores the relationship between environmental, social and governance (ESG) information and systemic risk, an increasingly important issue for both regulators and investors. While ESG ratings are widely used to assess a company's non-financial performance, the impact of these factors on financial stability and systemic risk is still under debate. By extending the Forecast Error Variance Decomposition (FEVD) method with a double regularization on both the underlying vector autoregressive (VAR) parameters and the covariance matrix of the VAR residuals, we are able to address the curse of dimensionality within each estimation. This allows us to examine how vulnerable a company is and how much systemic impact a company has given its specific ESG. Looking at a larger sample of European stocks over the period 2007–2022, we empirically show that both the best and worst ESG performers have the largest impact on the financial system in normal times. However, during a crisis, companies with the best ESG ratings generate significant spillovers throughout the system. These findings highlight the importance of incorporating ESG factors into systemic risk assessments and monitoring companies' ESG performance to ensure financial stability. Policymakers can benefit from this research by supporting investment in high ESG companies to mitigate relevant spillovers during stressed market conditions, when such companies are more interconnected.

1. Introduction

Accurate understanding of the factors that impact or hinder financial stability is of utmost importance, not only for regulators to foster stability, but also for investors and practitioners. Recently, investors have become increasingly interested in non-financial information, particularly regarding sustainability, to accurately gauge the performance and risk profiles of companies. This phenomenon has increased the relevance of environmental, social and governance (ESG) ratings, which are employed to evaluate a company's non-financial performance and the degree of transparency reporting (Refinitiv, 2023). This is especially important in regards to the Corporate sustainability reporting which entered force in the beginning of 2023 (European Commission, 2023). Despite ongoing debate about the relationship between ESG information and corporate performance and risk, ESG scores are commonly determined by various rating providers, including Bloomberg and Reuters, using a range of criteria, measurements, and quantitative and qualitative methods. These scores are typically ranged between 0 and 100, with higher scores indicating more responsible ESG behavior, and are often grouped into rating classes (e.g., A, B, C, D) based on predefined thresholds or quartiles established from the ESG score values. Relevant sources on this topic include the studies by Bhattacharya and Sharma (2019) and Berg and Lange (2020).

Research on the impact of ESG factors on financial stability and systemic risk is still limited and, as the special issue by Battiston et al. (2021) shows, many open questions and research gaps remain. However, some recent studies found that companies with higher ESG scores generally have lower financial distress and default risk (Boubaker et al., 2020) and contribute less to systemic risk (Eratalay and Cortés Ángel, 2022). Furthermore, by providing a tool to assess the exposure of a given portfolio to transition risk, Alessi et al. (2021) showed that, in times of stress, when greener and more transparent companies outperform brown stocks, losses would be incurred at the global level. In addition, ESG funds may be more resilient to contagion in periods of lower volatility (Cerqueti et al., 2021). In the banking sector, higher ESG scores are associated with a lower probability of sanctions for Italian banks, and lower environmental costs and risks may help promote financial stability during crises (Murè et al., 2021). In addition, according to Kanas et al. (2023), banks are negatively related to CO2 emissions, suggesting the development of a lower carbon economy in this sector. While previous research has shown that ESG data can help capture the systemic risk of different firms (Bax et al., 2022), questions

* Corresponding author.
E-mail addresses: karoline.bax@tum.de (K. Bax), giovanni.bonaccolto@unikore.it (G. Bonaccolto), sandra.paterlini@unitn.it (S. Paterlini).

https://doi.org/10.1016/j.jfs.2024.101221
Received 15 April 2023; Received in revised form 12 January 2024; Accepted 22 January 2024
Available online 2 February 2024

about spillovers remain unanswered. However, recently, Iqbal et al. (2022) showed that negative returns transmitting more strongly, especially during the COVID-19 crisis. Additionally, Chen and Lin (2022) find that extreme spillovers exhibit asymmetric characteristics and are more significant than median spillovers, with the Northern American and European markets being the main risk transmitters. Furthermore, socially responsible stock markets possess lower risk than conventional ones, with varying contributions to systemic risk across regions and market phases, particularly during financial crises (Ameur et al., 2020).

To gain further insights into the impact of ESG factors on financial stability and systemic risk, we leverage on the Forecast Error Variance Decomposition (FEVD) method introduced by Diebold and Yılmaz (2014), which has been widely cited and employed in the financial literature (Alter and Beyer, 2014; Apostolakis and Papadopoulos, 2015; Chevallier et al., 2018; Xu et al., 2019; Bostanci and Yilmaz, 2020; Iwanicz-Drozdowska et al., 2021; Andrieş et al., 2022; Greenwood-Nimmo et al., 2023; Pham et al., 2023). We estimate the spillovers among a large set of companies, and assess whether their systemic impact and vulnerability degree are affected by their ESG score. We adopt regularization techniques to deal with high-dimensional problems, where the number of variables is large compared to the sample size. Specifically, on the one hand, we resort to the approach proposed by Demirer et al. (2017) and Gross and Siklos (2019), who penalized the parameters of the underlying Vector Autoregressive (VAR) model, using the Least Absolute Shrinkage and Selection Operator (LASSO) and the Elastic Net (ELNET) introduced by Tibshirani (1996) and Zou and Hastie (2005), respectively. On the other hand, we point out the fact that the FEVD estimation also depends on the covariance matrix of the VAR residuals. Therefore, we need to address the curse of dimensionality when estimating such covariance matrix, in addition to the VAR coefficients. For this purpose, we correct the ill-conditioned sample estimator typically adopted in the FEVD estimation using the sparse estimator introduced by Rothman (2012).

As a result, from a methodological viewpoint, we extend the FEVD estimation by proposing a double regularization on both the VAR parameters and the covariance matrix of the VAR residuals. To the best of our knowledge, this is the first study which estimates the FEVD model building on this double regularization.

From an empirical point of view, we find important results. The relationship between both the systemic impact and the degree of vulnerability of companies and their ESG score is not obvious. It is time-dependent, and the evidence from stable periods is significantly different from that observed during tail events. Indeed, systemic impact and vulnerability tend to be increasing functions of ESG scores during stressed market conditions. In contrast, there is a kind of U-shape during stable periods as previously investigated by Bax et al. (2022). These findings allow us to highlight important policy implications. Shocks to firms characterized by higher ESG scores could have relevant spillovers throughout the system, especially during extreme adverse events such as financial crises, when the stronger co-movements among firms and the associated risk of contagion threaten the stability of the entire economy. The take-home message for policymakers is that encouraging investment in shares of companies that are more responsible from an ESG perspective could be an important policy measure to prevent or mitigate relevant spillovers during stressed market regimes when such companies are more interconnected.

We enrich our empirical analysis by comparing our method with alternative approaches in terms of selection process, validation and out-of-sample predictive accuracy, evaluating the stability of the resulting solutions. Moreover, we show that our main findings are confirmed by using different datasets, characterized by a larger cross section and obtained from a different data provider. We also find interesting results by considering potential size effects and by conducting an industry-level analysis.

The paper is structured as follows. Section 2 provides the details on the methodology we propose to estimate the regularized FEVD model. Section 3 describes the data and the implementation choices for our methodology. We report and discuss the empirical results in Section 4, and develop a rich set of robustness checks in Section 5. Finally, Section 6 concludes the paper.

2. Regularized forecast error variance decomposition

Let $\boldsymbol{y}_t = \left[y_{1,t} \cdots y_{N,t}\right]'$ be an $N \times 1$ vector, the entries of which are the returns of N stocks observed at time t, with $t = 1, \ldots, T$. We study the presence, propagation and dynamics of spillovers among these stocks by employing the Forecast Error Variance Decomposition (FEVD) method (Diebold and Yılmaz, 2014), which builds on the estimation of the following covariance-stationary Vector Autoregressive (VAR) model:

$$\boldsymbol{y}_t = \boldsymbol{\nu} + \sum_{k=1}^{p} \boldsymbol{\phi}_k \boldsymbol{y}_{t-k} + \boldsymbol{\epsilon}_t, \tag{1}$$

where $\boldsymbol{\phi}_k$ is an $N \times N$ matrix of slope parameters, $\boldsymbol{\nu} = \left[\nu_1 \cdots \nu_N\right]'$ is an $N \times 1$ intercept vector and $\boldsymbol{\epsilon}_t \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$ is the vector of errors, with $\mathbb{E}(\boldsymbol{\epsilon}_t \boldsymbol{\epsilon}_s') = 0$ and $s \neq t$ (Pesaran and Shin, 1998; Lütkepohl, 2007).

The covariance-stationary VAR model given in Eq. (1) can be rewritten as the infinite moving average representation:

$$\boldsymbol{y}_t = \boldsymbol{\mu} + \sum_{k=0}^{\infty} \boldsymbol{\Psi}_k \boldsymbol{\epsilon}_{t-k}, \tag{2}$$

where $\boldsymbol{\mu}$ is the unconditional expected value of $\boldsymbol{y}_t$, and the $N \times N$ parameter matrix $\boldsymbol{\Psi}_k$ is a function of $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p$, being obtained from the following recursive relationship:

$$\boldsymbol{\Psi}_k = \boldsymbol{\phi}_1 \boldsymbol{\Psi}_{k-1} + \boldsymbol{\phi}_2 \boldsymbol{\Psi}_{k-2} + \cdots + \boldsymbol{\phi}_p \boldsymbol{\Psi}_{k-p}, \tag{3}$$

with $\boldsymbol{\Psi}_0 = \boldsymbol{I}_N$ and $\boldsymbol{\Psi}_k = \mathbf{0}$ for $k < 0$, where $\boldsymbol{I}_N$ is the $N \times N$ identity matrix, whereas $\mathbf{0}$ is a zero matrix (Pesaran and Shin, 1998).

We use the generalized impulse response function proposed by Pesaran and Shin (1998), that, in contrast to the orthogonalized impulse responses, is not affected by the ordering of the variables in the underlying VAR model. Specifically, the entry placed on the ith row and jth column of the h-step generalized variance decomposition matrix is defined as follows (Pesaran and Shin, 1998; Diebold and Yılmaz, 2014):

$$\theta_{i,j}^{g}(h) = \frac{\sigma_{jj}^{-1} \sum_{l=0}^{h} \left(\boldsymbol{e}_i' \boldsymbol{\Psi}_l \boldsymbol{\Sigma} \boldsymbol{e}_j\right)^2}{\sum_{l=0}^{h} \left(\boldsymbol{e}_i' \boldsymbol{\Psi}_l \boldsymbol{\Sigma} \boldsymbol{\Psi}_l' \boldsymbol{e}_i\right)}, \tag{4}$$

where 'g' stands for 'generalized', σ_{jj} is the jth diagonal element of $\boldsymbol{\Sigma}$, and $\boldsymbol{e}_j$ is an $N \times 1$ selection vector with one as entry j and zeros elsewhere.

$\theta_{i,j}^{g}(h)$ given in Eq. (4) is the proportion of the h-step ahead forecast error variance of variable i which is accounted for by the innovations in variable j, with $i, j = 1, \ldots, N$. $\theta_{i,j}^{g}(h)$ depends on: (i) $\boldsymbol{\Psi}_l$, which, in turn, is a function of $\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p$; and (ii) $\boldsymbol{\Sigma}$: the covariance matrix of $\boldsymbol{\epsilon}_t$.

$\boldsymbol{\phi}_1, \ldots, \boldsymbol{\phi}_p$ are typically estimated by implementing the Ordinary Least Squares (OLS) method on each of the N equations of the VAR model defined in Eq. (1); see, among others, Lütkepohl (2007). Let $\boldsymbol{\phi}_{k,j}$ be the jth row of $\boldsymbol{\phi}_k$, we define the vector $\boldsymbol{\beta}_j = \left[\nu_j \ \boldsymbol{\phi}_{1,j} \cdots \boldsymbol{\phi}_{p,j}\right]'$, which includes all the parameters specific to equation j of the VAR model given in Eq. (1), where the response variable is $y_{j,t}$, whereas the regressors are $\boldsymbol{y}_{t-1}, \ldots, \boldsymbol{y}_{t-p}$. The OLS estimate of $\boldsymbol{\beta}_j$ is then obtained as the solution of the following optimization problem:

$$\underset{\beta_j}{\operatorname{argmin}} \sum_{t=p+1}^{T} \left(y_{j,t} - \nu_j - \sum_{k=1}^{p} \boldsymbol{\phi}_{k,j} \boldsymbol{y}_{t-k} \right)^2,$$

for $j = 1, \ldots, N$.

In our study, we deal with high-dimensional problems in which $T < N$. In this framework, OLS estimates would exhibit high variance due to the large number of covariates compared to the sample size. For this reason, we introduce regularization techniques, that outperform

the OLS method by trading off a small increase in bias for a large decrease in variance (Bonaccolto et al., 2023). Other studies in the financial literature estimated FEVDs by regularizing the coefficients of the underlying VAR models. For instance, we refer to Demirer et al. (2017), who adopted the Least Absolute Shrinkage and Selection Operator (LASSO) introduced by Tibshirani (1996), and Gross and Siklos (2019), who used the Elastic Net (ELNET) introduced by Zou and Hastie (2005). We aim at identifying the most important companies which have the strongest impact on the entire network. For this purpose, we prefer LASSO to ELNET, as the former has a greater shrinkage impact and leads to sparser solutions. However, we take into account the fact that LASSO typically provides biased estimates, overshrinking the retained variables (Fan and Li, 2001). We address this issue using the post-LASSO method described below.

In a first step, we use LASSO to select the variables in $\boldsymbol{y}_{t-1},\dots,\boldsymbol{y}_{t-p}$ which have a relevant impact on $y_{j,t}$. For this purpose, we minimize the following loss function:

$$L\left(\boldsymbol{\beta}_j\right)=\sum_{t=p+1}^{T}\left(y_{j,t}-\nu_j-\sum_{k=1}^{p}\boldsymbol{\phi}_{k,j}\boldsymbol{y}_{t-k}\right)^2+\lambda_j\|\boldsymbol{\beta}_j\|_1, \tag{5}$$

where $\|\boldsymbol{\beta}_j\|_1$ is the sum of the absolute values of the entries of $\boldsymbol{\beta}_j$ (i.e. the ℓ_1-norm penalty), whereas $\lambda_j \geq 0$ is the tuning parameter which determines the intensity of the penalization: the greater λ_j, the greater the number of coefficients approaching zero, leading to sparser solutions (Hastie et al., 2009; Murphy, 2012).

We then identify those regressors which are LASSO-selected. Specifically, a regressor is LASSO-selected if its corresponding slope coefficient, resulting from the minimization of $L\left(\boldsymbol{\beta}_j\right)$ in Eq. (5), is different from zero. In a second step, we estimate the slope coefficients of the LASSO-selected covariates by minimizing again the loss function in Eq. (5), but now discarding the regressors which are not LASSO-selected and setting $\lambda_j = 0$. Note that the solution obtained in this second step coincides with the output of the standard OLS method implemented on a linear regression model in which the response variable is $y_{j,t}$ and the explanatory variables are the ones that are LASSO-selected in the first step. In contrast, the slope coefficients of the regressors which are not LASSO-selected in the first step are set equal to zero. We denote the post-LASSO estimate of $\boldsymbol{\beta}_j$ as $\widehat{\boldsymbol{\beta}}_j$, for $j = 1,\dots,N$.

After obtaining $\widehat{\boldsymbol{\beta}}_1,\dots,\widehat{\boldsymbol{\beta}}_N$, we aggregate these estimates by lag order, building the VAR coefficient matrices $\widehat{\boldsymbol{\phi}}_1,\dots,\widehat{\boldsymbol{\phi}}_p$. From such post-LASSO estimates, we then compute the residual vector of the VAR model in Eq. (1), denoted as $\widehat{\boldsymbol{\epsilon}}_t$. According to Belloni and Chernozhukov (2011) and Hautsch et al. (2014), among others, the post-LASSO estimation provides relevant improvements and outperforms the standard LASSO. Furthermore, we use $\widehat{\boldsymbol{\phi}}_1,\dots,\widehat{\boldsymbol{\phi}}_p$ to derive the estimate of $\boldsymbol{\Psi}_l$, denoted as $\widehat{\boldsymbol{\Psi}}_l$.

As said before, in addition to $\boldsymbol{\Psi}_l$, FEVD also depends on $\boldsymbol{\Sigma}$, which is typically estimated using the sample covariance matrix:

$$\boldsymbol{S}=\frac{1}{T-p}\sum_{t=p+1}^{T}\left(\widehat{\boldsymbol{\epsilon}}_t-\overline{\boldsymbol{\epsilon}}\right)\left(\widehat{\boldsymbol{\epsilon}}_t-\overline{\boldsymbol{\epsilon}}\right)', \tag{6}$$

where $\overline{\boldsymbol{\epsilon}}=(T-p)^{-1}\sum_{t=p+1}^{T}\widehat{\boldsymbol{\epsilon}}_t$.

Nevertheless, the sample estimator defined in Eq. (6) is typically unsatisfactory or ill-defined when N approaches to or is greater than $T-p$ (Meucci, 2005; Riccobello et al., 2022). For this reason, we introduce the sparse estimator of $\boldsymbol{\Sigma}$ proposed by Rothman (2012) in the statistical literature, that is positive definite and performs well in high-dimensional settings. Here, a LASSO-type penalty is adopted to encourage sparsity and a logarithmic barrier function is used to enforce positive definiteness. Specifically, following Rothman (2012) we propose the following correlation matrix estimator:

$$\widehat{\boldsymbol{\Gamma}}=\underset{\boldsymbol{\Gamma}\succ\mathbf{0}}{\operatorname{argmin}}\left(\|\boldsymbol{\Gamma}-\boldsymbol{R}\|_F^2/2-\tau\log|\boldsymbol{\Gamma}|+\delta|\boldsymbol{\Gamma}^-|_1\right), \tag{7}$$

where $\boldsymbol{\Gamma}\succ\mathbf{0}$ points out that $\boldsymbol{\Gamma}$ is symmetric and positive definite, $\boldsymbol{R}$ denotes the sample correlation matrix of $\widehat{\boldsymbol{\epsilon}}_t$, $\delta \geq 0$ is the tuning parameter which encourages sparsity, $\tau > 0$ is fixed at a small value, $|\boldsymbol{\Gamma}|$ is the determinant of $\boldsymbol{\Gamma}$, $\boldsymbol{\Gamma}^- = \boldsymbol{\Gamma} - \boldsymbol{\Gamma}^+$, where $\boldsymbol{\Gamma}^+$ is a diagonal matrix with the same diagonal as $\boldsymbol{\Gamma}$, $|\boldsymbol{\Gamma}^-|_q = \|\operatorname{vech}(\boldsymbol{\Gamma}^-)\|_q$ is the q-norm of the vector formed by stacking the columns of $\boldsymbol{\Gamma}^-$, $\|\boldsymbol{\Gamma}-\boldsymbol{R}\|_F$ is the Frobenius norm of $\boldsymbol{\Gamma}-\boldsymbol{R}$.

Building on the estimator given in Eq. (7), the covariance matrix estimator takes the following form:

$$\widehat{\boldsymbol{\Sigma}}=\left(\boldsymbol{S}^+\right)^{1/2}\widehat{\boldsymbol{\Gamma}}\left(\boldsymbol{S}^+\right)^{1/2}, \tag{8}$$

where $\boldsymbol{S}^+$ is a diagonal matrix with the same diagonal as $\boldsymbol{S}$.

As explained by Rothman (2012), regularizing on the correlation scale enables us to prove a faster convergence rate bound and produces a covariance estimator $\widehat{\boldsymbol{\Sigma}}$ that is invariant to scaling of the variables. To the best of our knowledge, this is the first study in which FEVD is estimated by regularizing both $\boldsymbol{\Psi}_l$ and $\boldsymbol{\Sigma}$. The estimate of $\theta_{i,j}^g(h)$ defined in Eq. (4) that we obtain using $\widehat{\boldsymbol{\Psi}}_l$ and $\widehat{\boldsymbol{\Sigma}}$ is denoted as $\widehat{\theta}_{i,j}^g(h)$. As highlighted by Diebold and Yılmaz (2014) and Gross and Siklos (2019), sums of forecast error variance contributions are not necessarily unity. Therefore, we normalize the sum of all row entries in the variance decomposition matrix as follows:

$$\widehat{\gamma}_{i,j}^g(h)=\frac{\widehat{\theta}_{i,j}^g(h)}{\sum_{j=1}^{N}\widehat{\theta}_{i,j}^g(h)}\cdot 100 \tag{9}$$

so that each $\widehat{\gamma}_{i,j}^g(h)$ ranges between 0 and 100, providing a quantitative measure for the pairwise directional connectedness from company j to company i (Gross and Siklos, 2019).

Following Diebold and Yılmaz (2014), we compute the following three indicators. First, the total directional connectedness from others to company i:

$$C_{i\leftarrow\bullet}^{(h)}=\frac{1}{N-1}\sum_{\substack{j=1\\ j\neq i}}^{N}\widehat{\gamma}_{i,j}^g(h), \tag{10}$$

which measures the vulnerability of company i.

Second, the total directional connectedness to others from company j:

$$C_{\bullet\leftarrow j}^{(h)}=\frac{1}{N-1}\sum_{\substack{i=1\\ i\neq j}}^{N}\widehat{\gamma}_{i,j}^g(h), \tag{11}$$

which measures the systemic impact of company j.

Third, the total connectedness or spillover index of the overall system:

$$C^{(h)}=\frac{1}{N(N-1)}\sum_{\substack{i,j=1\\ i\neq j}}^{N}\widehat{\gamma}_{i,j}^g(h). \tag{12}$$

3. Data description and empirical setup

In our work, we use the daily logarithmic return, daily market capitalization and annual environmental, social and governance (ESG) data, provided by Refinitiv, of $N = 294$ European (EU) companies. Specifically, these $N = 294$ companies are the constituents of the EURO STOXX 600 index for which we have data availability from January 2, 2007 to May 3, 2022. We classify the companies of our dataset into four ESG categories based on their ESG performance. Specifically, a company with an ESG score between 75 and 100 receives an A rating. Companies with a score between 50 and 75 have a B rating. C-rated companies are the ones with an ESG score between 25 and 50. Finally, we assign a D rating to the companies with a score below 25.

Our dataset spans a period of 16 years, and the ESG score of each company is provided with an annual frequency. In our empirical analysis, we perform both an annual and a full-sample analysis. In the first case, we estimate a given regularized variance decomposition matrix using the data taken from a single year. As a result, we obtain 16 different time-varying values of $\widehat{\gamma}_{i,j}^g(h)$ defined in Eq. (9), for $i, j = 1,\dots,294$.

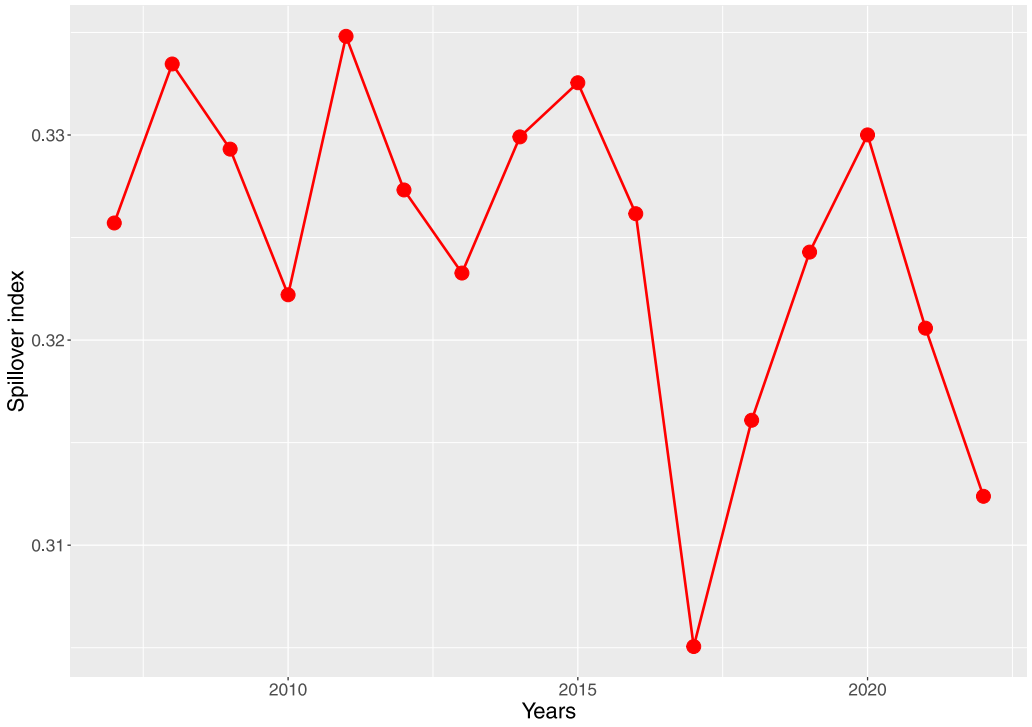


Fig. 1. Trend of the total spillover index $C_w^{(10)}$ from 2007 to 2022.

For each year, we cluster the 294 companies according to their ESG scores observed for that specific year. Therefore, the composition of such four clusters changes over time. In the second case, we estimate a unique regularized variance decomposition matrix using the full-sample dataset of daily returns. In this case, we compute the average ESG score of each company; that is, the mean of the 16 ESG scores observed for company j from 2007 to 2022, for $j = 1, \dots, 294$. We then obtain a 294×1 vector of average ESG scores, from which we cluster the companies into four classes, following the criterion described above. Specifically, Class A includes those companies with average ESG scores between 75 and 100. Class B (C) includes those companies with average ESG scores between 50 and 75 (25 and 50). In contrast, the companies belonging to class D have an average ESG score lower than 25.

As for the lag order of the VAR model defined in Eq. (1) and the FEVD temporal horizon in Eq. (4), we follow Gross and Siklos (2019) and set $p = 2$ and $h = 10$ in our empirical analysis. The optimal value of the tuning parameter λ_j in Eq. (5) is chosen by five-fold cross-validation, which is widely used in applied machine learning due to its good performance and flexibility; see, among others (Hastie et al., 2009; Murphy, 2012). Likewise, we select the optimal value of δ given in Eq. (7) using the tuning parameter selection method proposed by Rothman (2012), setting the number of random splits equal to five. As recommended by Rothman (2012), we set $\tau = 10^{-4}$ in Eq. (7), as this value leads to a stable solution of the optimization algorithm.

From the FEVD estimation, we obtain a directed and weighted network, that we analyze using the R package 'igraph'. The nodes of this network are the 294 EU companies in our dataset. The size of node j is proportional to $C_{\bullet \leftarrow j}^{(h)}$ defined in Eq. (11), which quantifies the systemic relevance of company j, for $j = 1, \dots, 294$. The width of the link which connects node j to node i is proportional to $\hat{\gamma}_{i,j}^{g}(h)$ defined in Eq. (9), for $i, j = 1, \dots, 294$ and $i \neq j$. We employ the Fruchterman and Reingold's (1991) algorithm to define the layout of the graph. This algorithm belongs to the class of force-directed algorithms, which typically use a physical simulation where some kind of attractive force are used to attract nodes connected by edges together. As a result, tightly connected clusters of nodes will show up close to each other, and those that are loosely connected will be repulsed towards the outside.

4. Empirical findings

In this section, we present the findings obtained from the empirical analysis. We initiate our discussion by examining the time-varying analysis in Section 4.1, wherein we showcase the estimates obtained for each specific year from 2007 to 2022. In Section 4.2, our focus shifts to the outcomes derived from the comprehensive full-sample analysis.

4.1. Time-varying estimation

In this section, we focus on the time-varying analysis. In particular, for each year from 2007 to 2022, we cluster the $N = 294$ EU companies of our dataset according to their year-specific ESG scores. By doing so, it could happen that a given company has a different ESG classification from year to year. Therefore, the results presented in this section allow us to check whether and to what extent the relationships between the systemic impact and the vulnerability degree of the 294 EU companies are affected by their time-varying ESG scores across different market regimes. The three indicators defined, respectively, in Eqs. (10)–(12) are then denoted as $C_{j \leftarrow \bullet, w}^{(10)}$, $C_{\bullet \leftarrow j, w}^{(10)}$ and $C_w^{(10)}$, for $j = 1, \dots, 294$ and $w = 2007, \dots, 2022$.

We first analyze the trend of the total spillover index $C_w^{(10)}$ in Fig. 1. Fig. 1 reflects the effects of important events, such as the subprime crisis and the lockdown due to the COVID-19 pandemic. Interestingly, the relevance of such tail events also emerges from the trend of the Composite Indicator of Systemic Stress (CISS) provided by the European Central Bank (see Fig. A.10 in Appendix). The greatest peak in Fig. 1 is observed in 2011, highlighting the impact of the EU sovereign debt crisis. In contrast, the effects of such event are less evident in Fig. A.10. Another important peak in Fig. 1 is observed in the year 2015, when the complicated negotiations between the Greek government and its international creditors threatened the risk of default and the potential exit from the monetary union.

We now assess the contribution of the different ESG classes to the overall spillover index. Starting from the systemic impact, for each year, we cluster the $C_{\bullet \leftarrow 1, w}^{(10)}, \dots, C_{\bullet \leftarrow 294, w}^{(10)}$ values into four groups: A, B, C and D. Specifically, $C_{\bullet \leftarrow j, w}^{(10)} \in \mathcal{R}$ if company j has ESG rating $\mathcal{R}$ in

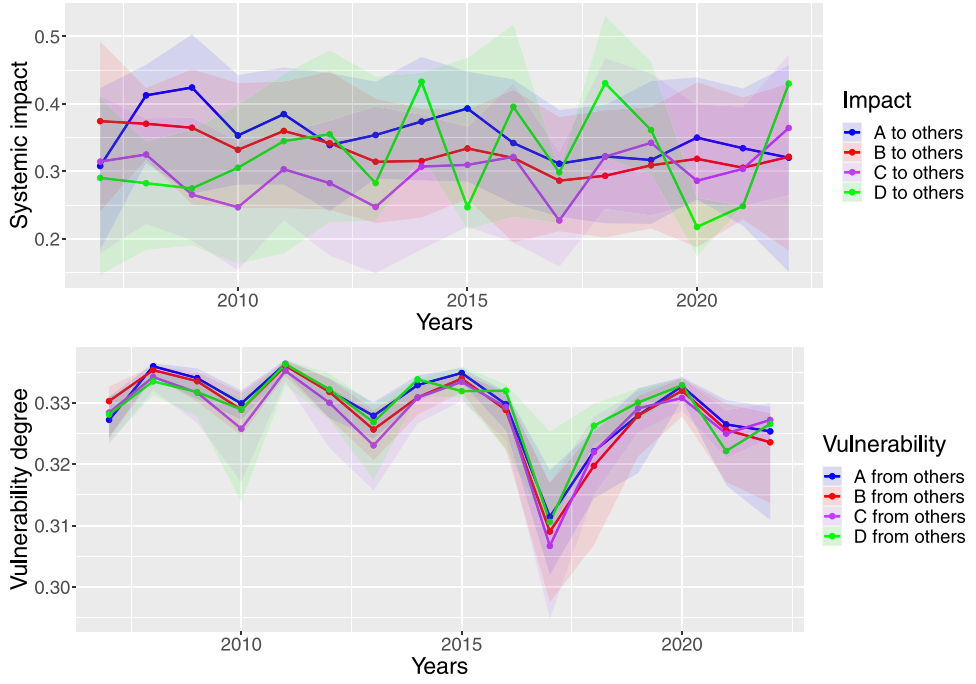


Fig. 2. Annual systemic impact and vulnerability degree clustered by ESG class.

year w, where $\mathcal{R}$ coincides with one of the four categories A, B, C or D. For each class and for each year, we then compute the first, second and third quartiles of the selected $C^{(10)}_{\bullet\leftarrow j,w}$ values. We display the results in the top panel of Fig. 2. Here, the solid lines reproduce the annual medians (or second quartile) of the four classes, whereas the ribbons are designed according to their respective first and third quartiles. Likewise, we cluster the vulnerability degrees $C^{(10)}_{1\leftarrow\bullet,w},\dots,C^{(10)}_{294\leftarrow\bullet,w}$ using a similar approach, and show the results in the bottom panel of Fig. 2. The A-rated companies tend to produce the greatest systemic impact in many years from 2007 to 2022. Interestingly, the D-rated companies have the greatest impact in six out of the 16 considered years, and their distance from the firms with a different rating becomes greater during stable periods, such as the years 2014 and 2018, when the CISS index approaches zero (see Fig. A.10). Another interesting result is the monotonic behavior from D to A in 2008, a year characterized by relevant stress according to the CISS index. The distances among the four ESG classes are less evident when looking at the vulnerability index (bottom panel of Fig. 2). In general, the ranking of the four ESG classes determined by the vulnerability degree is similar to the one determined by the systemic impact. For instance, we again observe the monotonic increasing trend from D to A in 2008.

The analysis of the indicators depicted in Fig. 2 can be further explored by identifying the composition of the other ESG classes towards which the systemic links in the top panel are directed, and the composition of the other ESG classes from which the vulnerability links in the bottom panel come from. Figs. 3 and 4 allow us to extend the analysis of such systemic and vulnerability links, respectively. We can see from Fig. 3 a clustering effect: the systemic impact of a given ESG class tends to be directed towards the same ESG class. This evidence is clearer in the top-left (bottom-left) panel of Fig. 3, where the A-rated (C-rated) companies have, on average, a greater impact to the companies of the same A (C) class. This clustering phenomenon is also present in the right panels of the same figure, in which we focus on the B- and D-rated companies, even if the distances among the four classes are less evident. The role of the A-rated companies emerges more clearly in Fig. 4. Here, we can see that the four ESG classes are particularly exposed to the impact of the A-rated companies. In some specific years, the greatest links come, on average, from the D-rated companies. In contrast, the C-rated companies tend to produce the lowest impact. The difference in the ranking between Figs. 3 and 4 is due to the fact that the rows of the variance decomposition matrix—which determine the vulnerability degrees—are normalized to sum up to one, whereas this is not true for the columns of the same matrix—which determine the systemic effects.

In addition to the analysis described above, we shed additional light on the relationships between systemic impact, vulnerability degree and ESG score with a regression approach. In particular, starting from the systemic impact, we estimate for each year w a regression model in which the values of the response variable are $C^{(10)}_{\bullet\leftarrow 1,w},\dots,C^{(10)}_{\bullet\leftarrow 294,w}$, while the values of the explanatory variable are the ESG scores specific to year w, for $w=2007,\dots,2022$. We estimate a local regression between the two variables of interest to emphasize the effects at the different regions of the ESG domain. Moreover, we make the trend of the estimation output clearer, cleaning it from the effect of noisy observations, by setting the span parameter of the local regression (which controls the degree of smoothing) equal to one. For each year from 2007 to 2022, Fig. 5 shows the fitted values (red lines) as well as the corresponding confidence intervals (shaded areas).

Fig. 5 highlights important results. First, the relationship between systemic impact and ESG score changes over time. We find an increasing trend during the years 2008 and 2009, which are characterized by the maximum level of distress according to Fig. A.10, due to the strong effects of the US subprime crisis. Therefore, during these years, the higher the ESG score is, the higher the systemic impact. In both years, the fitted value of the systemic impact ranges from a minimum value of around 0.25 corresponding to the minimum ESG score to a maximum value of around 0.40 corresponding to the maximum ESG score. The confidence bands are larger at the left and right tails of the ESG distribution, due to the relatively low number of companies characterized by low and high ESG scores during the years 2008 and 2009 (see Fig. A.11). This increasing trend is also observed in 2010 when focusing on the medium-high ESG values; in contrast, the curve is quite flat at low ESG values. Afterwards, in the following years, this increasing trend disappears, leaving room for the U shape discussed above, which becomes clear during the year of maximum stability (i.e. 2014; see Fig. A.10). After intermediate years in which the curve is quite flat, we again observe, mainly at medium-high ESG values, an increasing trend in 2020: the year of the outbreak of the COVID-19

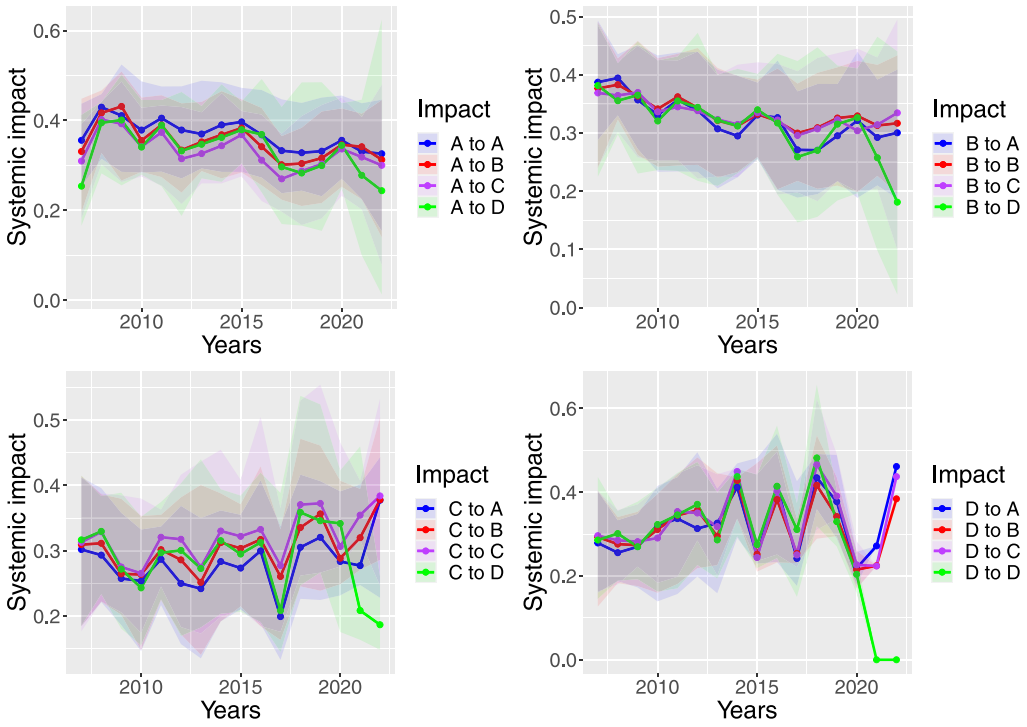


Fig. 3. Systemic impact of the A, B, C and D ESG classes to the A, B, C and D ESG classes.

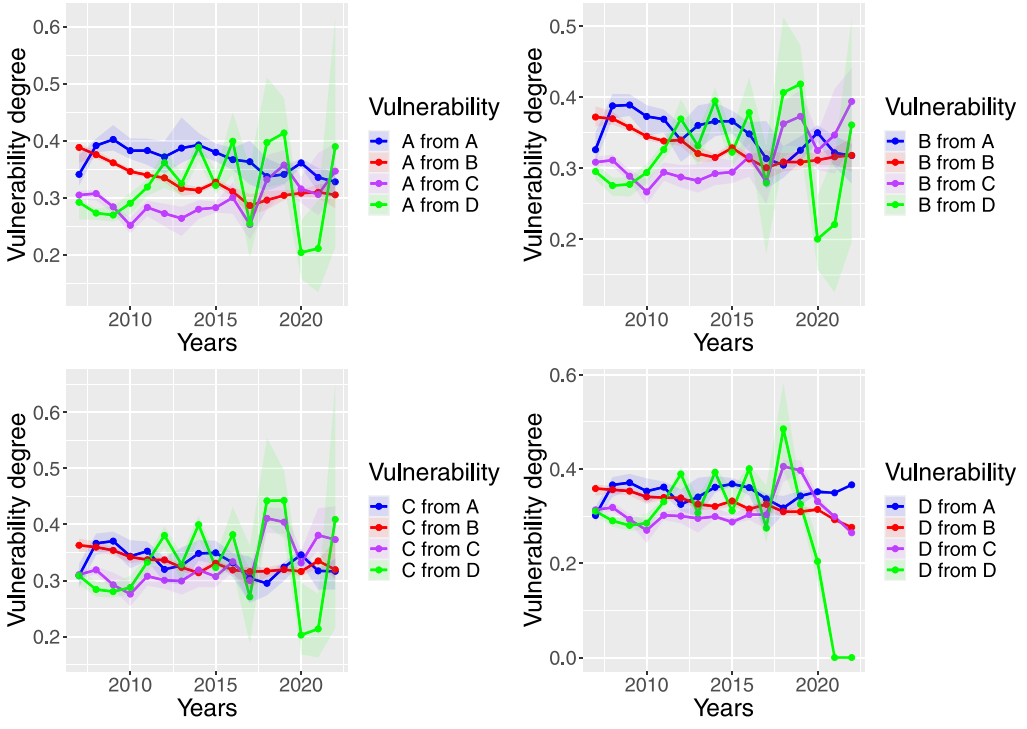


Fig. 4. Vulnerability degree of the A, B, C and D ESG classes from the A, B, C and D ESG classes.

pandemic, the effects of which are also clear in Fig. A.10. Interestingly, from 2007 to 2022, the confidence bands become larger at low ESG values and tighter at high ESG values, reflecting the fact that the ESG scores of the 294 companies in our dataset tend to increase over time. Therefore, a greater uncertainty in the estimates is associated to a lower number of observations, and vice versa.

We now focus on the relationship between the ESG scores and the vulnerability degree. We estimate the same regression model discussed above, with the difference that the values of the response variables are now the $C^{(10)}_{1\leftarrow\bullet,w}, \ldots, C^{(10)}_{294\leftarrow\bullet,w}$ indicators. We show the resulting estimates in Fig. 6. Interestingly, we find concave curves from 2008 to 2011. Here, the vulnerability degree is an increasing function of the ESG score, mainly at low and medium ESG values.

The curves are quite flat and become slightly decreasing at high ESG values. However, the vulnerability degree corresponding to high ESG scores is significantly greater than the one associated to low ESG levels.

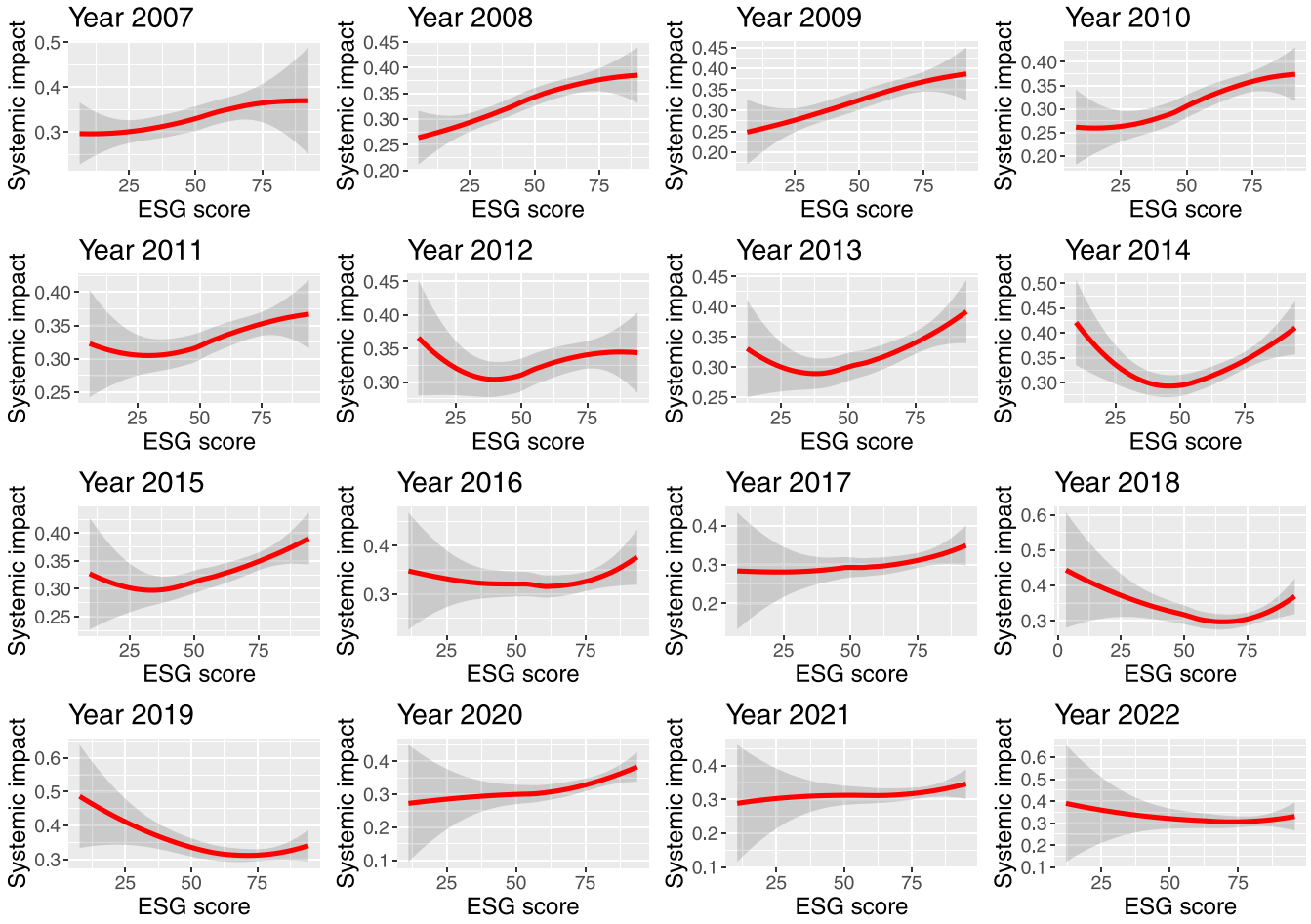


Fig. 5. Systemic impact defined by the $C^{(10)}_{\bullet\leftarrow j,w}$ indicator as a local function of the ESG score, for each year from 2007 to 2022 (the confidence interval is delimited by the shaded area). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

In contrast to Fig. 5, where the curve of the year 2011 is more similar to the U-shaped curves of the following years, the fitted values in Fig. 6 in the year 2011—a year characterized by market turmoils due to the EU sovereign debt crisis—follow the same trend of the estimates obtained during the time interval 2008—2010. As a result, if an increasing trend is mainly associated to stressed periods, while U-shaped curves are obtained during stable phases of the markets, it means that the effects of the EU sovereign debt crisis emerge when looking at the vulnerability of the 294 EU companies. From the year 2012, there is a U shape, which is particularly evident in 2014, similar to the evidence drawn from Fig. 5. Again, the curves become quite flat in the most recent years.

All in all, the results discussed in this section point out that the relationships between ESG score, systemic impact and vulnerability degree change over time. We find a clear distinction between stressed periods, where the systemic impact and the vulnerability degree are increasing function of the ESG scores, and stable periods, where the U shape is predominant. As a result, shocks to A-rated companies could produce relevant spillovers throughout the overall system, especially during tail events, such as financial crises. As highlighted by Bonaccolto et al. (2019), among others, the co-movements among firms increase during such periods, and the risk of contagion threatens the stability of the entire economy due to more relevant spillover effects.

4.2. Full-sample analysis

We continue the empirical analysis by describing the results obtained from the full-sample data. This additional analysis provides an overall picture of the connections among the 294 EU companies described in Section 3 derived from a long time interval, which spans 16 years, characterized by the occurrence of relevant tail events as well as stable periods (see Fig. A.10).

We display in Fig. 7 the network of the 294 companies clustered by average ESG score. The size of the nodes indicates the relative impact. The four panels of Fig. 7 show the same network. However, they emphasize with the red color the role and impact of the different ESG classes: A (Fig. 7(a)), B (Fig. 7(b)), C (Fig. 7(c)) and D (Fig. 7(d)). In each panel, the nodes representing the companies belonging to a given ESG class, as well as the links that start from them, are depicted in red, whereas the remaining nodes and links are depicted in gray.

We can see from Fig. 7 that most companies belong to the A and B ESG classes. Specifically, classes A, B, C and D include, respectively, 71, 158, 59 and 6 companies. This evidence is due to the fact that companies improved their average ESG score from 2007 to 2022 as shown in Fig. A.11 given in Appendix. This behavior can be either due to an improvement in the ESG dimensions or due to an increased awareness of more disclosure data. For this reason, the red links in Figs. 7(a) and 7(b) play a dominant role. However, we highlight an important difference between Figs. 7(a) and 7(b). In Fig. 7(a), A-rated companies are predominantly located in the central part of the network, and their outgoing links are mainly directed towards other central nodes highlighting their importance. In contrast, B-rated companies are scattered throughout the network, and their presence and impact are more evident in the periphery of the same network. Indeed, the central region of the network given in Fig. 7(b) presents a more evident gray

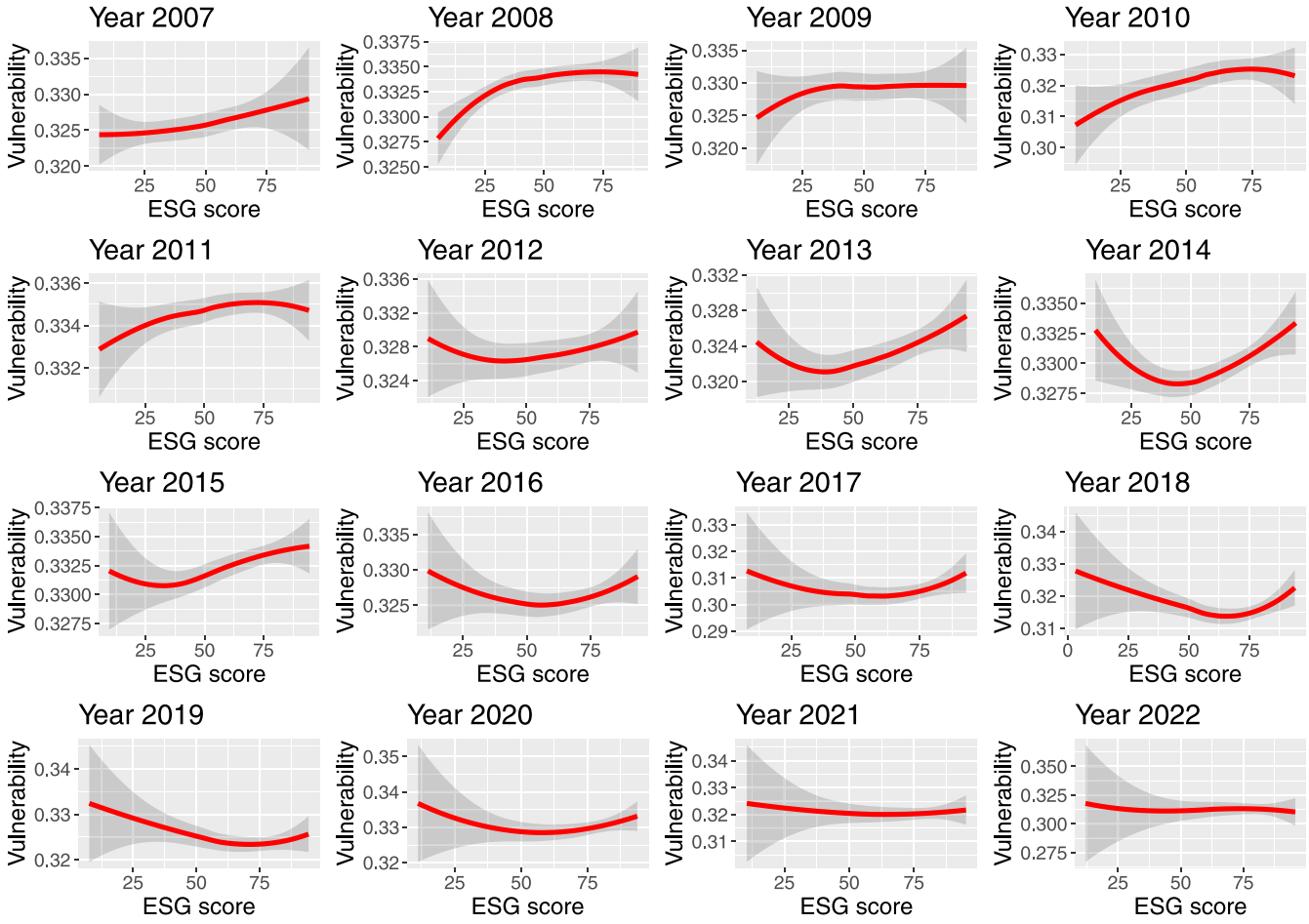


Fig. 6. Vulnerability defined by the $C^{(10)}_{i\leftarrow\bullet,w}$ indicator as a local function of the ESG score, for each year from 2007 to 2022 (the confidence interval is delimited by the shaded area).

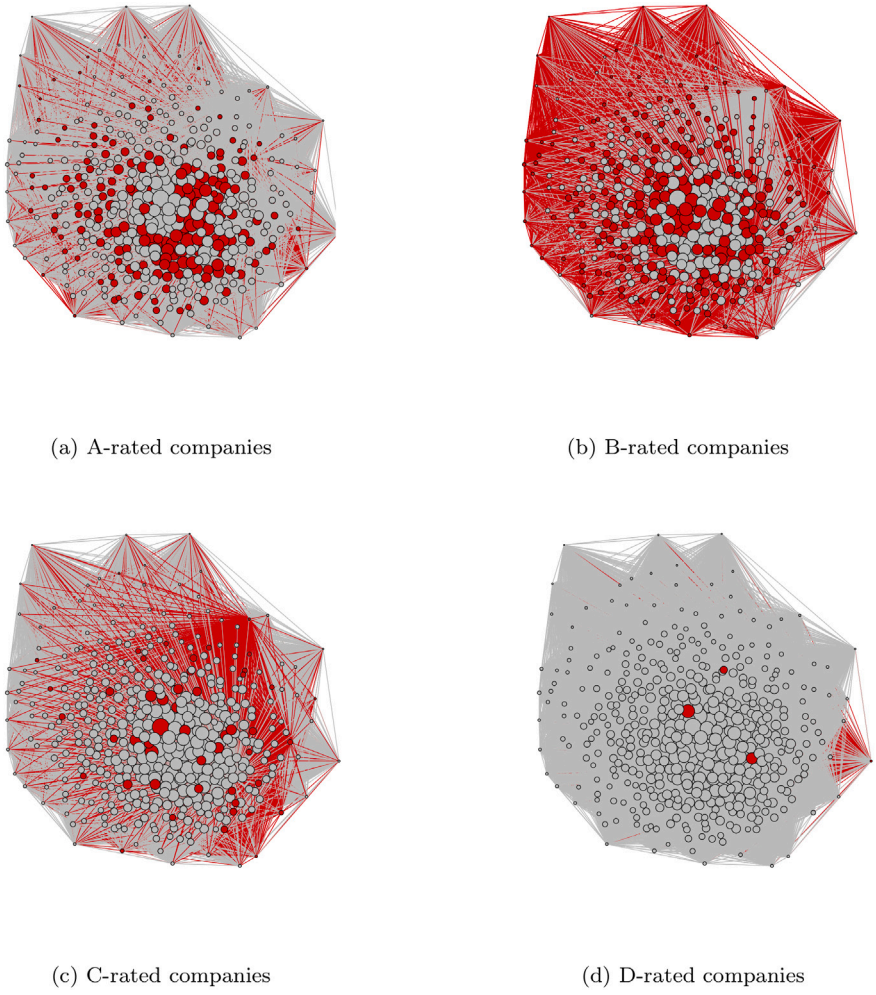
area with respect to Fig. 7(a), despite the number of B-rated companies is significantly larger than the number of A-rated firms.

Despite there are only six D-rated companies, they play an important role (see Fig. 7(d)). Indeed, four of them are placed in the central part of the network, and their size (reflecting their systemic impact) is noticeable compared to the other nodes of the graph. In general, we observe the same phenomenon in the overall network: the greater the size of nodes, the more central they are within the graph. Therefore, systemically important firms tend to be more interconnected and show greater impact. Finally, we examine the role of C-rated companies in Fig. 7(c). These companies are scattered throughout the network, with a relatively small presence in its central region. A few of them have a noticeable size and their outgoing links are mainly directed towards peripheral nodes.

The analysis described above reveals important information about the distribution of the nodes and links within the overall network. Nevertheless, it is difficult to evaluate the strength or magnitude of such links as well as the distinction between ingoing and outgoing links from Fig. 7. We then pair Fig. 7 with Fig. 8, in which we show the boxplots of the $C^{(10)}_{i\leftarrow\bullet}$ and $C^{(10)}_{\bullet\leftarrow j}$ values, in cyan and red, respectively, clustering the 294 nodes into the four ESG classes. We remind the reader that $C^{(10)}_{i\leftarrow\bullet}$ and $C^{(10)}_{\bullet\leftarrow j}$ reflect, respectively, the vulnerable and systemic degrees of the nodes given in Fig. 7, summing up the edge weights of the adjacent edges for each vertex.

We can observe in Fig. 8 an interesting U-shape distribution of the weighted degrees across the ESG classes. The D-rated companies tend to take greater weighted degrees, followed by the A-rated companies. This is an interesting result which supports the evidence provided by Fig. 7. That is, despite the number of A- and D-rated companies is significantly lower compared to the B- and C-rated companies (71 + 6 = 77 versus 158 + 59 = 217), the firms belonging to the two extreme classes (A and D) tend to have a greater impact on the overall network. Indeed, A- and D-rated companies are mainly positioned in the central area of the network displayed in Fig. 7, being more interconnected with the other nodes of the graph, especially the ones characterized by a greater size. Despite the B-rated companies are the most numerous, providing a larger number of links, their impact is relatively small. Therefore, such companies have many interconnections, whose strength is, however, limited. We then find a clear coherence between Figs. 7 and 8: stronger links are mainly concentrated in the central area of the network, where the presence of B- and C-rated companies is relatively limited. Another interesting evidence provided by Fig. 8 is that each of the four ESG classes presents similar distributions of systemic and vulnerable weighted degrees. A greater difference between the distributions of $C^{(10)}_{i\leftarrow\bullet}$ and $C^{(10)}_{\bullet\leftarrow j}$ is observed for the D-rated companies.

Summing up the results discussed in this section, the A- and D-rated companies exhibit a greater impact within the estimated network, with a U-shape distribution across the different ESG classes. As highlighted by Bax et al. (2022), A- and D-rated firms might be attracting more attention from investors due to the positive and negative screening investment policies, respectively: best in class and worst in class are two investment criteria commonly used to decide which assets to focus on. The role of the A-rated companies becomes more critical, as



(a) A-rated companies (b) B-rated companies

(c) C-rated companies (d) D-rated companies

Fig. 7. The role of the different classes of companies (highlighted in red), clustered by average ESG score, within the network estimated from the full-sample data observed from January 2, 2007 to May 3, 2022. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

they are more numerous than D-rated companies.[1] Indeed, in a world characterized by greater attention to sustainability issues, companies keep improving their ESG performance. A-rated companies are then subject to a greater pressure, as investors have great expectations for them. The take-home message for EU policymakers is that they should be aware that not only shocks to D-rated firm but also to A-rated firms could produce intensive spillovers throughout the overall system, due to their increasing presence as well as to their strong degree of interconnections.

5. Robustness analysis

In this section, we present the results of seven robustness checks. In Section 5.1, we compare the method employed in our study with alternative well-known approaches in terms of selection process. We assess the stability of the selection process in Section 5.2, whereas Section 5.3 focuses on the validation procedure. Section 5.4 takes into account potential size effects. We present the results obtained from a dataset characterized by a larger cross section in Section 5.5. In Section 5.6, we develop an industry-level analysis. Finally, we study the results derived from the ESG scores and the return time series obtained from a different data provider in Section 5.7.

[1] This result is consistent with the findings obtained from the time-varying analysis described in Section 4.1. Moreover, the relevance of the A-rated companies becomes more evident when conducting a sector analysis (see Section 5.6).

5.1. Comparison with alternative methods

An important input of the regularized FEVD presented in our study is $\boldsymbol{\Psi}_l$ in Eq. (4), which, in turn, depends on $\boldsymbol{\phi}_k$ given in the VAR model defined in Eq. (1). We estimate $\boldsymbol{\phi}_k$ by employing the post-LASSO method. That is, in a first step, we LASSO-select the relevant covariates of each VAR equation. In a second step, we estimate the slope parameters of the selected variables. In contrast, we set the slope coefficients of the regressors which are not LASSO-selected equal to zero. As a result, the selection made by LASSO in the first step plays a critical role, as it determines the sparsity of the final solutions and the magnitude of the spillover index.

In our study, we focus on LASSO. However, other alternative approaches are available in the literature. The most known could be classified into Mutual Information based methods, wrapper methods, sequential selection algorithms, classifiers, and embedded methods. Detailed information about such methods are provided by Langley (1994), Kohavi and John (1997), Guyon et al. (2002), Guyon and Elisseeff (2003), Alpaydin (2004), Law et al. (2004), Ding and Peng (2005), Chuang et al. (2008) and Lazar et al. (2012). Therefore, it is interesting to assess the similarity of the solutions provided by LASSO and other alternative methods. In our study, we first consider the Minimum Redundancy-Maximum Relevance (MRMR) approach introduced by Ding and Peng (2005). MRMR offers an efficient selection of relevant and non-redundant features, ranking them by maximizing the mutual information with the response variable and minimizing the

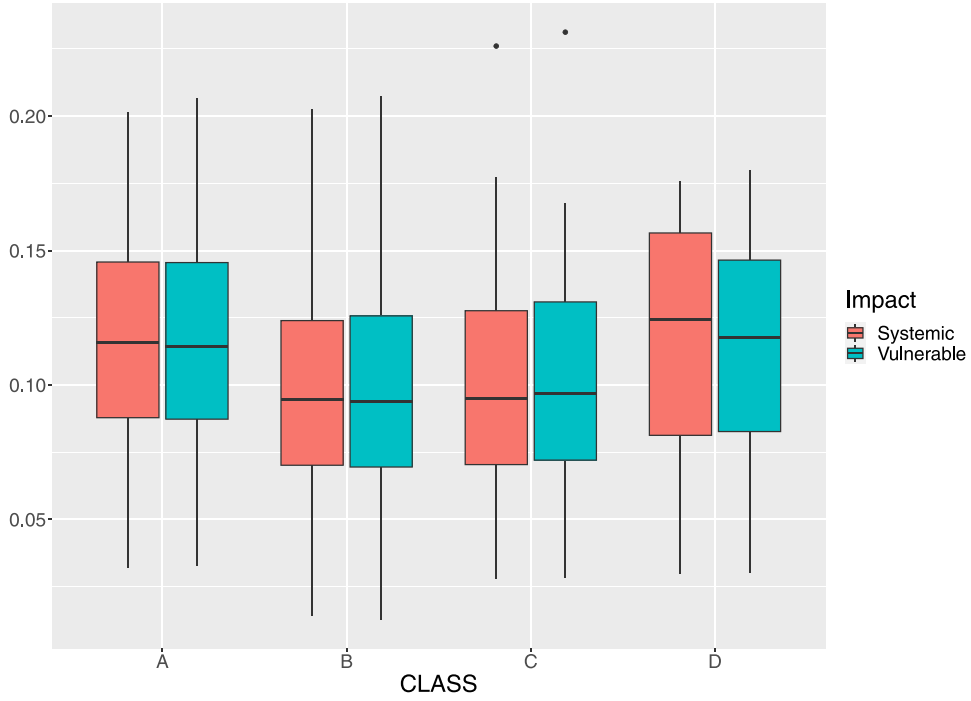


Fig. 8. Distribution of the $C^{(h)}_{\bullet\leftarrow j}$ (systemic weighted degree) and $C^{(h)}_{i\leftarrow\bullet}$ (vulnerable weighted degree) indicators defined in Eqs. (11) and (10), respectively, clustered by ESG class. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

average mutual information with all the covariates which are selected in previous iterations (Jay et al., 2013). Second, we also employ the Random Forest (RF) method introduced by Breiman (2001). RF offers substantial gains in classification and regression accuracy by using ensembles of trees. The final results are then obtained by aggregating over the ensemble (Biau, 2012).

We compare LASSO, MRMR and RF for each equation of the VAR model in (1). This exercise is expensive in terms of computational time. Therefore, for simplicity, we set the lag order p in Eq. (1) equal to one when implementing the three alternative approaches, so that their comparison is made under the same conditions. We use the full-sample data employed in Section 4.2. Focusing on equation j of the VAR model, we employ LASSO and build a 294×1 vector of binary data, denoted as D_j^{LASSO}, that is defined as follows. Element i of D_j^{LASSO} takes the value of one if regressor i is LASSO-selected in equation j, and the value of zero otherwise, with $i, j = 1, \dots, 294$. We denote as NS_j the sum of the entries of D_j^{LASSO} (i.e. the number of selected regressors in equation j), with $j = 1, \dots, 294$.

We also obtain the ranking provided by MRMR of the 294 regressors in equation j of the VAR model, from 1 (that corresponds to the most relevant regressor) to 294 (that corresponds to the least relevant regressor). We then build a 294×1 vector of binary data, denoted as D_j^{MRMR}, where element i takes the value of one if its ranking score is less than or equal to NS_j, and the value of zero otherwise. We repeat this procedure for $j = 1, \dots, N$, then identifying the top-ranked regressors for all 294 equations of the VAR model. Likewise, we build the vectors $D_1^{\text{RF}}, \dots, D_{294}^{\text{RF}}$ building on the ranking provided by RF.

The vectors D_j^{LASSO}, D_j^{MRMR} and D_j^{RF} allow us to identify the selected or top-ranked regressors of equation j, for $j = 1, \dots, 294$. We evaluate their similarity by employing the Tanimoto and Simple Matching measures described by Härdle and Simar (2019). These measures range from zero (i.e. zero similarity) to one (i.e. maximum similarity). They are specifically designed to assess the similarity between two objects having a binary structure and well capture the impact of the co-occurrences of both the 0 and 1 values. We display in Fig. 9 the boxplots of the Tanimoto and Simple Matching values, computed for each pair of alternative methods and for all $N = 294$ equations of the VAR model.

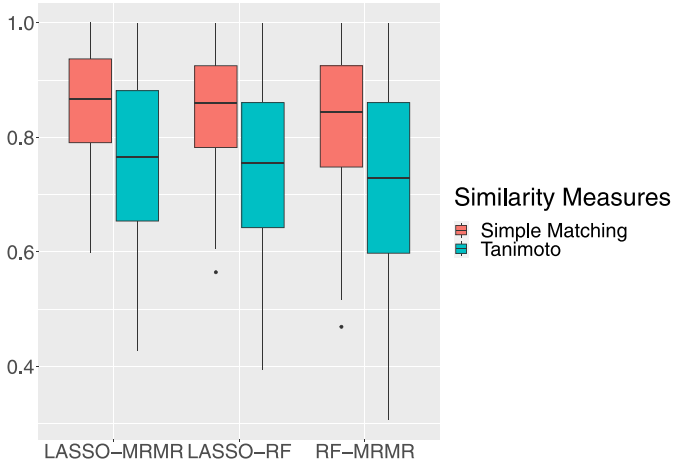


Fig. 9. Similarity between the LASSO, MRMR and RF methods.

In general, we find high similarities between the three methods. On average, the Simple Matching measure leads to a greater similarity. The distributions across the pairs of alternative methods are similar for both the Simple Matching and Tanimoto measures.

5.2. Stability assessment

Another important point, strictly related to the one discussed in Section 5.1, concerns the stability of the selection made by the method employed in our study: LASSO. In this context, the subsampling-based strategy turns out to be very useful and informative. Building on this strategy, we generate a large number of subsamples from the entire population, and then compute a measure of robustness (Lazar et al., 2012). As explained in Section 3, our dataset spans a long period from January 2, 2007 to May 3, 2022, characterized by stable and stressed phases of the financial markets. Therefore, it is interesting to

dynamically evaluate the impact of such changes of regimes on the stability of the selection made by LASSO. We can conduct this dynamic evaluation using a rolling window scheme, which allows us to generate a large number of rolling subsamples.

Specifically, we define the first subsample as the one which spans the time periods $t = 1, \dots, 1000$. From time to time, we remove the oldest observations and add the new ones, so that we keep constant the sample size. Therefore, the second subsample includes the time periods $t = 2, \dots, 1001$, and so on until we use our entire dataset. Overall, we obtain 3002 rolling subsamples. For simplicity, we estimate our VAR model with a lag order equal to one, as done in Section 5.1.

Following Lazar et al. (2012), for each equation of the VAR model, we evaluate the overall stability of the selection made by LASSO by computing the average over all pair-wise comparisons between different solutions:

$$R_{tot,j} = \frac{2\sum_{s=1}^{3002}\sum_{z=s+1}^{3002} R(f_{j,s}, f_{j,z})}{3002 \cdot 3001}, \tag{13}$$

where $f_{j,s}$ is the outcome of LASSO applied to the jth VAR equation and subsample s, and $R(f_{j,s}, f_{j,z})$ is a similarity measure between $f_{j,s}$ and $f_{j,z}$ (Lazar et al., 2012), for $j = 1, \dots, 294$.

We create a connection with Section 5.1 by defining $f_{j,s}$ as $D_{j,s}^{\text{LASSO}}$, which is a 294×1 vector of binary data, where element i is equal to one if regressor i is LASSO-selected in equation j and subsample s, and the value of zero otherwise. Again, we use the Tanimoto and Simple Matching similarity measures to be consistent with the analysis developed in Section 5.1. We display the values of $R_{tot,1}, \dots, R_{tot,294}$ defined in Eq. (13) in the top (Tanimoto) and bottom (Simple Matching) panels of Fig. A.12 given in Appendix. The high values displayed in Fig. A.12 point out the stability in the selection made by LASSO. Focusing on the Tanimoto similarity measure, the minimum and maximum values of $R_{tot,j}$ ($j = 1, \dots, 294$) are equal to 0.66 and 1.00, respectively, with first and third quartiles of 0.93 and 0.98. When considering the Simple Matching similarity measure, the minimum and maximum values of $R_{tot,j}$ ($j = 1, \dots, 294$) are equal to 0.79 and 1.00, respectively, with first and third quartiles of 0.96 and 0.99. In both cases, the medians take high values: 0.96 and 0.98, respectively.

5.3. Validation

The choice of the tuning parameter λ_j in Eq. (5) plays a crucial role in our study. Indeed, it determines the variables that are selected in the first step of our post-LASSO method and, hence, the sparsity of the final solutions. The parameters of the LASSO-selected variables are then estimated in a subsequent second step, and could be used to provide predictions from the VAR model. Therefore, the choice of λ_j has an important impact on both the model selection and the out-of-sample performance prediction. As highlighted by Guyon and Elisseeff (2003), one should first distinguish the problem of model selection from that of evaluating the final performance of the predictor. For that last purpose, it is important to set aside an independent test set. The remaining data is used both for training and performing model selection.

In our study, the model selection depends on the value of λ_j in Eq. (5). We determine the value of this parameter using the five-fold cross-validation (5F-CV) method. Other approaches are available in the literature. For instance, we mention the leave-one-out procedure, which, however, typically provides estimates affected by high variance and is expensive in terms of computational time (Vapnik, 2006; Guyon and Elisseeff, 2003; Hastie et al., 2009). Another possibility is to increase the number of folds, so that we can consider ten-fold cross-validation (10F-CV), which is also widely used in the literature (Hastie et al., 2009). Likewise, Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) are also well-known methods; see, among others, Hurvich (1985), Hastie et al. (2009) and Torri et al. (2018).

In this section, we compare the out-of-sample performance of 5F-CV, 10F-CV, AIC and BIC in terms of prediction accuracy. Building on the same window size of 1000 observations employed in Section 5.2, we consider a first subsample which spans the time periods $t = 1, \dots, 1000$. From this subsample, we define the first training set starting at day $t = 1$ and ending at day $t = 980$, from which we conduct different model selections based on 5F-CV, 10F-CV, AIC and BIC, respectively. Again, similar to Sections 5.1 and 5.2, we implement LASSO by setting $p = 1$ in Eq. (5) for computational simplicity. For each method taken separately and keeping the data of the first training set, we then estimate the final slope coefficients of the LASSO-selected variables in the second step of our post-LASSO method. Building on the parameters estimated from the first training set, we then predict the values of the response variable from $t = 981$ to $t = 1000$, so that we use the overall information of the first subsample. By doing so, as recommended by Guyon and Elisseeff (2003), we separate the first training set ($t = 1, \dots, 980$) from the first validation set ($t = 981, \dots, 1000$). From the first validation set, we then compute, for each competing method, the mean squared error (MSE). We repeat this procedure with three additional non-overlapping training and validation sets. The second training and validation sets span the time intervals $t = 1001, \dots, 1980$ and $t = 1981, \dots, 2000$, respectively. The third training and validation sets include the time intervals $t = 2001, \dots, 2980$ and $t = 2981, \dots, 3000$, respectively. Finally, the fourth training and validation sets span the time intervals $t = 3001, \dots, 3980$ and $t = 3981, \dots, 4000$, respectively. For each method, we then aggregate the MSE values computed from the four validation sets.

We display the boxplots of the MSE values resulting from the $N = 294$ VAR equations in Fig. A.13 given in Appendix. 10F-CV, 5F-CV and BIC provide a similar out-of-sample performance across the $N = 294$ VAR equations, whereas AIC leads to the worst predictions. On average, 5F-CV slightly outperforms the competing approaches, with a mean value of 0.0263%, followed by 10F-CV (0.0266%), BIC (0.0269%) and AIC (0.0337%).

5.4. Size effect

In this section, we take into account the possibility that the relationship between the systemic impact and the ESG scores could be affected by a latent size effect. That is, the possibility that companies with larger size invest more in variables linked to responsibility and sustainability issues and hence have a higher ESG score. Thus, these companies would be categorized into the better rated (A-rated) cluster. As a result, the real driver of the systemic impact would be the size of companies rather than their ESG score. We shed light on this point by estimating, for each year from 2007 to 2022, a standard regression model which includes both the ESG score and size as regressors, whereas the response variable is the systemic impact. Specifically, the values of the response variables are the systemic indicators $C_{\bullet\leftarrow 1,w}^{(10)}, \dots, C_{\bullet\leftarrow 294,w}^{(10)}$, where w denotes the reference year, for $w = 2007, \dots, 2022$. As for the regressors, we highlight the fact that the ESG score and the size level have a different scale. We then standardize, for each specific year, their values so that the magnitude of their coefficients is comparable. We denote the values of the first regressor as $\text{ESG}_{1,w}^{\star}, \dots, \text{ESG}_{294,w}^{\star}$, which are obtained from the standardization of the original $\text{ESG}_{1,w}, \dots, \text{ESG}_{294,w}$ scores. As for the second regressor (i.e. size), we compute the mean of the daily market capitalization values specific to company j and year w, denoted as $\text{MC}_{j,w}$, for $j = 1, \dots, 294$ and $w = 2007, \dots, 2022$. We then obtain $\text{MC}_{1,w}^{\star}, \dots, \text{MC}_{294,w}^{\star}$: the standardized values of $\text{MC}_{1,w}, \dots, \text{MC}_{294,w}$.

For each $w = 2007, \dots, 2022$, we then estimate the following regression model:

$$C_{\bullet\leftarrow j,w}^{(10)} = \alpha_0 + \alpha_1 \text{ESG}_{j,w}^{\star} + \alpha_2 \text{MC}_{j,w}^{\star} + \eta_{j,w}, \tag{14}$$

where $\eta_{j,w}$ is the error term, whereas α_0, α_1 and α_2 are scalar parameters, whose OLS estimates are denoted as $\widehat{\alpha}_0$, $\widehat{\alpha}_1$ and $\widehat{\alpha}_2$, respectively.

We report the OLS estimates of the parameters given in Eq. (14), along with their p-values, in Table 1. These results highlights at least two interesting points. First, $\text{ESG}_{j,w}^{\star}$ is statistically significant at the 5% level in nine out of the overall 16 years. Interestingly, these years

Table 1
Estimation of the regression model defined in Eq. (14).

Year	ESG		Size		$\lvert\widehat{\alpha}_1\rvert/\lvert\widehat{\alpha}_2\rvert$
	$\widehat{\alpha}_1$ (%)	p-value (%)	$\widehat{\alpha}_2$ (%)	p-value (%)	
2007	***2.4468***	***0.8486***	−0.5652	54.0872	4.3287
2008	***3.4631***	***0.0000***	−0.5880	32.7637	5.8902
2009	***3.6766***	***0.0004***	−0.6797	38.5581	5.4092
2010	***3.7288***	***0.0008***	0.4415	59.0371	8.4464
2011	***1.8298***	***0.8719***	1.1969	8.5166	1.5288
2012	1.0979	17.8746	0.7804	33.8800	1.4068
2013	***2.4005***	***0.1513***	0.1502	84.1279	15.9783
2014	***1.7546***	***1.4791***	0.3878	58.8292	4.5248
2015	***2.3583***	***0.0374***	0.7302	26.5855	3.2295
2016	0.4954	54.8387	1.4343	8.2969	0.3454
2017	1.3754	8.5722	1.3254	9.7686	1.0378
2018	−0.4476	58.0405	0.4373	58.9173	1.0236
2019	−1.4947	5.3162	1.0532	17.2373	1.4192
2020	***2.0948***	***1.1275***	0.3054	71.0304	6.8591
2021	0.5350	46.8794	1.1144	13.1871	0.4801
2022	−0.4220	68.6564	1.6129	12.3700	0.2616

Notes: For each year from 2007 to 2022 (first column), the table reports the results of the estimation of the regression model in which the response variable is the systemic impact of the companies in our dataset, whereas the regressors are the standardized ESG scores (ESG) and the standardized size (SIZE) of the same companies. The bold values indicate the significant values at a 5% significance level.

include the US subprime crisis, the EU sovereign debt crisis and the outbreak of the COVID-19 pandemic, supporting the evidence found in Section 4.1: the relationships between the ESG score and the systemic impact become more relevant during tail events. Second, $\mathrm{MC}^{\star}_{j,w}$ is never significant at the 5% level. Therefore, the ESG score is the relevant driver of the systemic impact during particular time periods, net of the size effect. This evidence is supported by the ratio $|\widehat{\alpha}_1|/|\widehat{\alpha}_2|$, which is almost always greater than one, pointing out a greater impact, in absolute value, of the ESG score on the systemic impact with respect to the size of the 294 companies.

5.5. Increasing the cross section dimension

We develop the empirical analysis in Sections 4.1 and 4.2 building on a cross section of $N = 294$ EU companies: the constituents of the EURO STOXX 600 index for which we have data availability from January 2, 2007 to May 3, 2022. It is now interesting to replicate this analysis by considering a shorter time interval, which provides the availability of a larger number of companies. Specifically, we now consider a new dataset which spans the trading days from January 4, 2016 to December 31, 2021, thereby allowing us to include a larger number of EU companies, equal to $N = 454$. By doing so, on the one hand, we exclude important events, such as the 2007—2009 US subprime crisis and the EU sovereign debt crisis. However, on the other hand, we are also aware that, after the Paris Agreement and the introduction of the EU Directive for reporting non financial information, the availability of ESG scores increases. Moreover, a larger cross section allows us to perform an accurate industry-level analysis (see Section 5.6).

We display in Fig. A.14 reported in Appendix the trend of the overall spillover index obtained from the time-varying estimates, with $N = 454$. Similar to Fig. 1, the effects of the COVID-19 pandemic are also evident in Fig. A.14.

Again, the most systemically relevant companies are the ones with ratings A and D. The A-rated companies exhibit the highest systemic impact in 2016, a particular year which comes immediately after the adoption of the Paris Agreement in 2015. In the subsequent years, the D-rated companies take the lead in producing the most significant systematic influence (see Fig. A.15 in Appendix). Similar to the findings discussed in Section 4.1, the distances among the four ESG classes are less evident when looking at the vulnerability index (bottom panel of Fig. A.15).

Once more, we find an evident clustering effect (Fig. A.16). From Fig. 3, we had already observed that the systemic impact exerted by a specific ESG class predominantly targets companies within the same ESG class. This pattern is also pronounced in the top-left panel of Fig. A.16, where the A-rated companies, on average, exert the most influence on companies of their own A rating. This effect is also shown strong for the years 2018–2020 for the D-rated companies in line with the impact shown in Fig. A.15. Furthermore, this clustering tendency is also discernible in the other panels, where, however, the distinctions between the classes are weaker. In Fig. A.17, one can see that all four ESG classes are impacted by the D-rated companies, as shown by their larger vulnerability degree. This is especially pronounced starting from year 2017.

Following the same approach described in Section 4.1, and starting from the systemic impact, we now model, for each year w (with $w = 2016, \dots, 2021$), a local regression where $C^{(10)}_{\bullet\leftarrow 1,w}, \dots, C^{(10)}_{\bullet\leftarrow 454,w}$ are the response variable values, whereas the explanatory variable values are the ESG scores observed in year w. We display the estimates derived from this regression in Fig. A.18 given in Appendix. We again find that the relationship between the response and the explanatory variables is not constant, changing over time. For every year, we see a U-shape in Fig. A.18, which is also present in many panels of Fig. 5. We find a similar evidence by comparing Figs. 6 and A.19.

We now focus on the full-sample analysis. Similar to Figs. 7, A.20 given in Appendix shows the (larger) network of the $N = 454$ companies of our new dataset clustered by average ESG score. Here, we have 171 A-rated companies, 231 B-rated companies, 48 C-rated companies and only 4 D-rated companies. The distribution of the $N = 454$ companies across these four classes highlights their improvements in terms of ESG scores over time. Fig. A.20 corroborates our prior findings. The distinctions between Figs. A.20(a) and A.20(b) remain consistent. In Fig. A.20(a), companies with rating A are centralized within the network, with their outgoing links predominantly connecting to other central nodes, underscoring their pivotal role. Conversely, the B-rated companies disperse more broadly across the network, with a pronounced influence and footprint on its periphery.

We then refer to Fig. A.21 where we present the boxplots of $C^{(10)}_{i\leftarrow\bullet}$ and $C^{(10)}_{\bullet\leftarrow j}$ in cyan and red, respectively. The 454 assets are grouped into four ESG classes. We remind the reader that $C^{(10)}_{i\leftarrow\bullet}$ and $C^{(10)}_{\bullet\leftarrow j}$ denote the vulnerability and systemic degrees of nodes, respectively. Similar to the U-shape we already observed in Fig. 8, we find a similar magnitude, order, and shape in Fig. A.21. Once again, we observe a minimal difference between the systemic and vulnerable weighted degrees. The largest difference is noticed among D-rated companies. Both A- and D-rated companies are predominantly positioned in the central area of the network, as depicted in Fig. A.20. Notably, even though the number of D-rated companies is much smaller than that of A-rated companies, their presence in the central region is significant. On the other hand, B- and C-rated companies display the least impact, even if they have a large number of connections. This reaffirms our earlier observation that stronger links are mainly found in the network's central area, where B- and C-rated companies are comparatively scarce.

5.6. Industry-level analysis

The larger cross section ($N = 454$) considered in Section 5.5 allows us to conduct an accurate industry-level analysis. In particular, we estimate the regularized FEVD for each individual economic sector. The 454 companies of our dataset are classified into 10 sectors: basic materials (49 firms), consumer cyclicals (63 firms), consumer non-cyclicals (40 firms), energy (18 firms), financials (87 firms), healthcare (32 firms), industrials (81 firms), real estate (20 firms), technology (40 firms) and utilities (24 firms). We show in Fig. A.22 the weighted degrees resulting from each sector. It is interesting to note that all four ESG classes appear in only two sectors: consumer cyclicals and financials. Class C is absent in the energy sector, whereas class D does

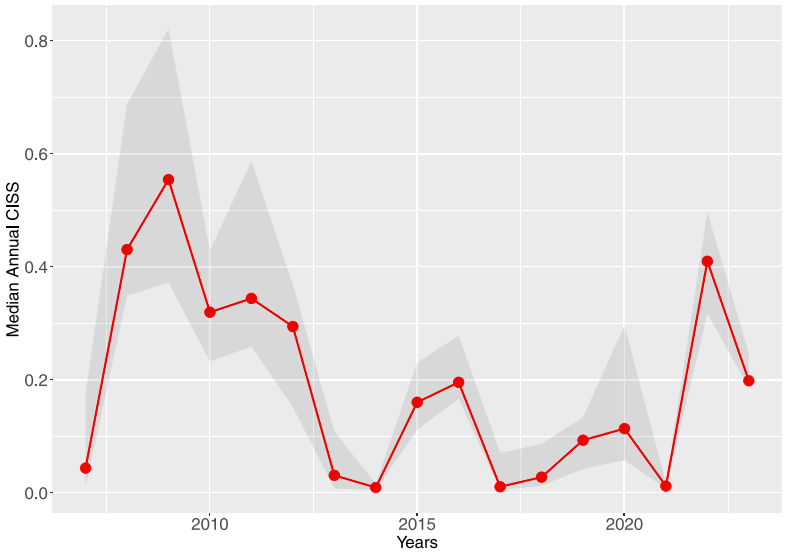


Fig. A.10. Annual median (red line) of the Composite Indicator of Systemic Stress provided (with a daily frequency) by the European Central Bank (the shaded area is delimited by the first and third quartiles, computed for each year, of the original daily values).

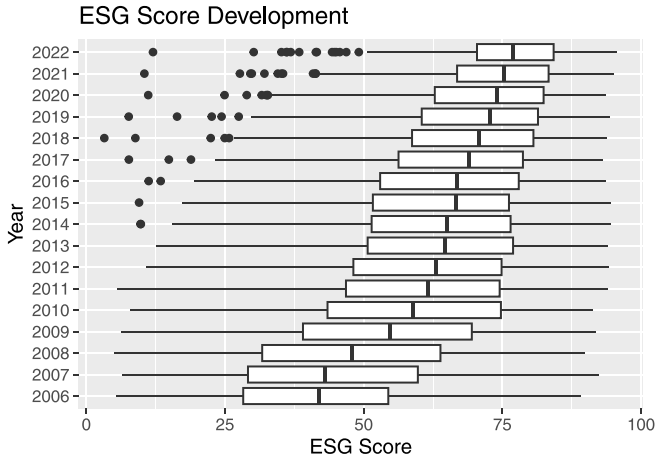


Fig. A.11. ESG development over time.

not appear in the remaining sectors. The A-rated companies have the greatest impact in the majority of the analyzed sectors. The D-rated companies exhibit the largest $C^{(h)}_{\bullet\leftarrow j}$ and $C^{(h)}_{i\leftarrow\bullet}$ values within the consumer cyclicals sector, followed by the A-rated companies, providing then a U shape similar to the one showed in Fig. 8. This is the only case in which we observe the U shape. Indeed, there is a decreasing trend from class A to class D in the energy and financials sectors. In general, this decreasing trend appears in many other sectors.

5.7. Results with a different data provider

We are aware that the disagreement across different data providers on ESG is a well-known issue. Therefore, it is interesting to compare the main findings discussed above with the outcome resulting from the return time series and ESG ratings obtained from a different data provider. More specifically, we employ the Bloomberg ESG score, which is founded on Bloomberg's perspective on ESG financial materiality. This score functions as a weighted generalized mean (power mean) of Pillar Scores, where the respective weights are ascertained by the pillar priority ranking. The values within this scoring model range from 0 to 10, with 10 being the most favorable score. Specifically, we classify the companies of this new dataset into three different classes: (i) the class of companies with rating CCC (which corresponds to a ESG score lower than or equal to 1.429); the class of companies with rating from B to BBB (which includes the ESG scores greater than 1.429 and lower than or equal to 5.714); and (iii) the class of companies with rating from A to AAA (which includes the ESG scores greater than 5.714). We still focus on the ESG scores and return times of the constituents of the EURO STOXX 600 index (observed from January 4, 2016 to December 31, 2021), the number of which is equal to $N = 469$.

For simplicity, we focus on the full-sample analysis.[2] We display the estimated network in Fig. A.23 given in Appendix. Similar to Fig. 7, the companies with more extreme rating scores have a central position within the network in Fig. A.23, whereas the ones in the middle class are scattered throughout the same graph.

Fig. A.24 in Appendix displays the boxplots of the $C^{(10)}_{i\leftarrow\bullet}$ and $C^{(10)}_{\bullet\leftarrow j}$ values, in cyan and red, respectively, clustering the companies of the new dataset into the three ESG classes. We remind the reader that $C^{(10)}_{i\leftarrow\bullet}$ and $C^{(10)}_{\bullet\leftarrow j}$ reflect, respectively, the vulnerable and systemic degrees of the nodes given in Fig. A.23. Interestingly, we observe from Fig. A.24 a great systemic relevance of companies with the highest ESG rating, supporting the industry-level findings discussed in Section 5.6.

6. Concluding remarks

This paper provides evidence of the importance of considering ESG factors in systemic risk assessment while also being the first to estimate the FEVD model based on a double regularization. We handle high-dimensional problems by penalizing the parameters of the VAR model while also correcting the ill-conditioned sample estimator typically used in FEVD estimation with a sparse estimator.

We find evidence that the relationship between both the systemic impact and the degree of vulnerability of companies and their ESG score is time-dependent and large differences between stable periods during tail events exists. Overall, systemic impact and vulnerability tend to be increasing functions of ESG scores during stressed market conditions. Furthermore, we show that firms with the best and worst ESG ratings have the highest vulnerability and potential systemic impact in normal times. Moreover, we show that the best and worst ESG performers can have a significant impact on the financial system in

[2] The results obtained from the time-varying estimates are available upon request.

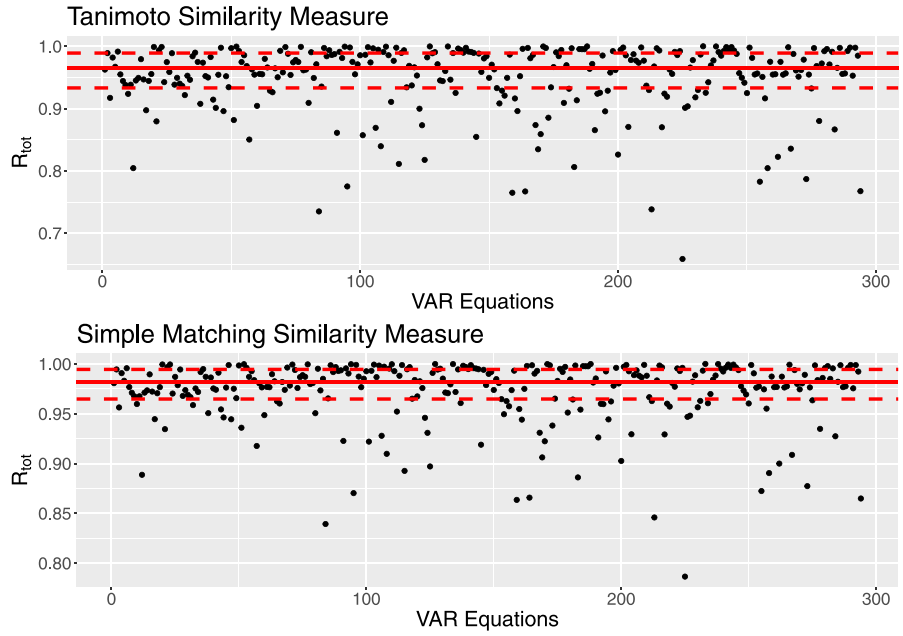


Fig. A.12. R_{tot} values computed on each of the $N = 294$ equations of the VAR model, using the Tanimoto and Simple Matching similarity measures. *Notes*: the red solid and dashed lines represent the median, first and third quartiles of the R_{tot} values displayed in each panel.

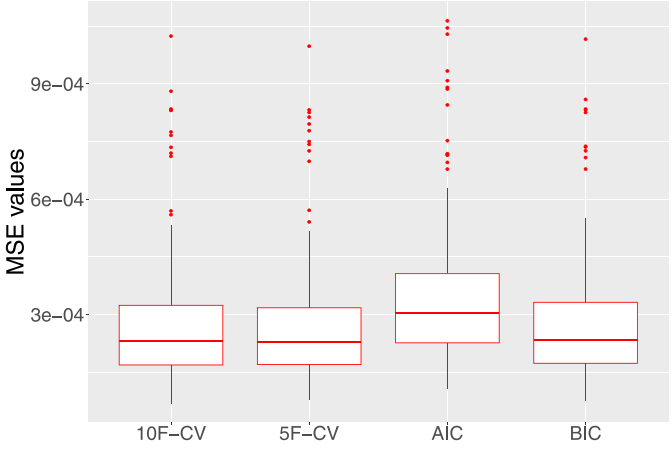


Fig. A.13. Mean squared errors provided by four different validation methods: ten-fold cross-validation (10F-CV), five-fold cross-validation (5F-CV), Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

normal times, while companies with the best ESG ratings have larger spillover effects throughout the system during crises. Our main findings are supported by a rich robustness analysis, in which we compare our selection process with alternative approaches, assess its stability across a large set of rolling subsamples and test the impact of our validation method in terms of predictive accuracy. These findings are consistent with the ones derived from different datasets, characterized by a larger cross section and obtained from a different data provider. We also find interesting results by considering size effects as well as by conducting an industry-level analysis.

Overall, this study provides important insights into the role of ESG factors in systemic risk and highlights the need for policymakers, investors and other stakeholders to monitor and integrate ESG information into their decision-making processes, especially during extreme tail events such as financial crises, when strong co-movement across firms exists and the associated risk of contagion threaten the stability of the entire economy. Policymakers should encourage investment into companies with high ESG scores to prevent or mitigate relevant spillovers during stressed market regimes when such companies are more interconnected. By doing so, they can help promote financial stability and mitigate potential risks to the financial system.

In conclusion, this research underlines the importance of integrating ESG information into decision-making processes in order to promote financial stability and reduce potential risks to the system as a whole.

CRediT authorship contribution statement

Karoline Bax: Conceptualization, Data curation, Methodology, Visualization, Writing – original draft, Writing – review & editing. **Giovanni Bonaccolto:** Conceptualization, Data curation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Sandra Paterlini:** Conceptualization, Data curation, Methodology, Writing – original draft, Writing – review & editing.

Acknowledgments

We sincerely thank the editor and the two anonymous referees for their insightful and valuable feedback. Furthermore, we extend our thanks to Marina Toepfer and all the participants of the 2023 Workshop on Current Research Trends in Sustainable Finance and of the International Conference on Computational and Financial Econometrics (CFE 2023) for their invaluable input and constructive discussions. Sandra Paterlini acknowledges financial support from COST Action HiTEc, CA21163.

Appendix. Additional figures

See Figs. A.10–A.24.

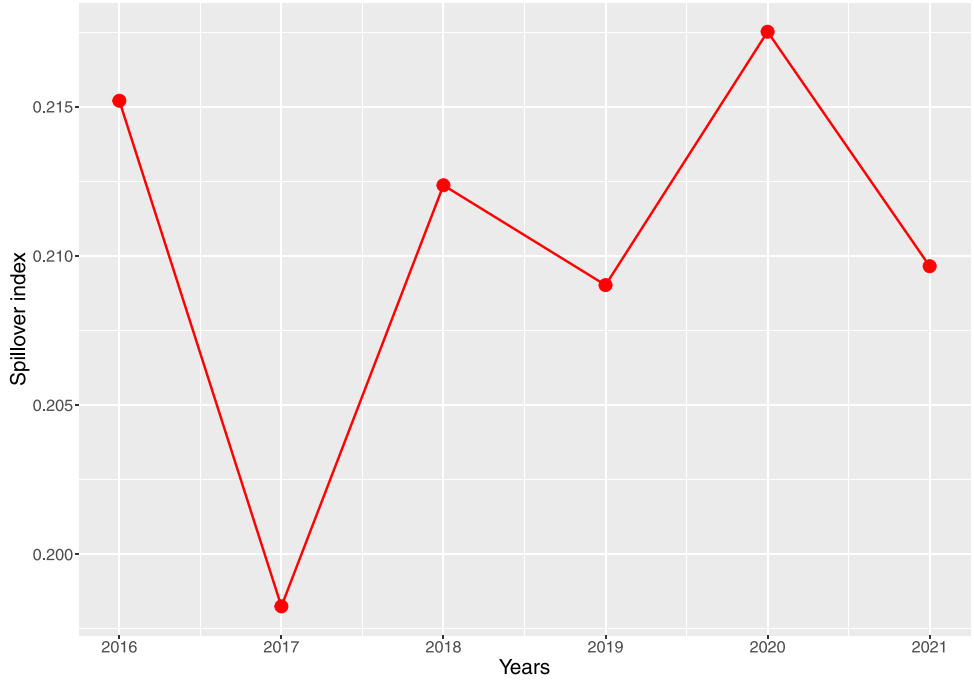


Fig. A.14. Trend of the total spillover index $C_w^{(10)}$ estimated from our dataset which includes $N = 454$ companies and spans the time interval from January 4, 2016 to December 31, 2021.

Fig. A.15. Annual systemic impact and vulnerability degree clustered by ESG class, estimated from our dataset which includes $N = 454$ companies and spans the time interval from January 4, 2016 to December 31, 2021.

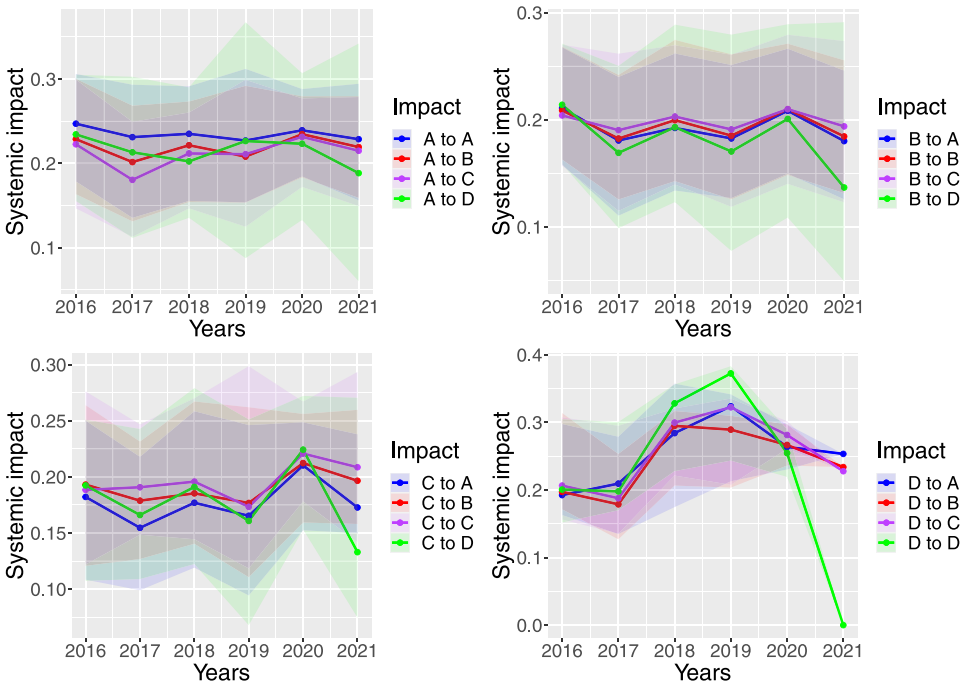


Fig. A.16. Systemic impact of the A, B, C and D ESG classes to the A, B, C and D ESG classes, estimated from our dataset which includes $N = 454$ companies and spans the time interval from January 4, 2016 to December 31, 2021.

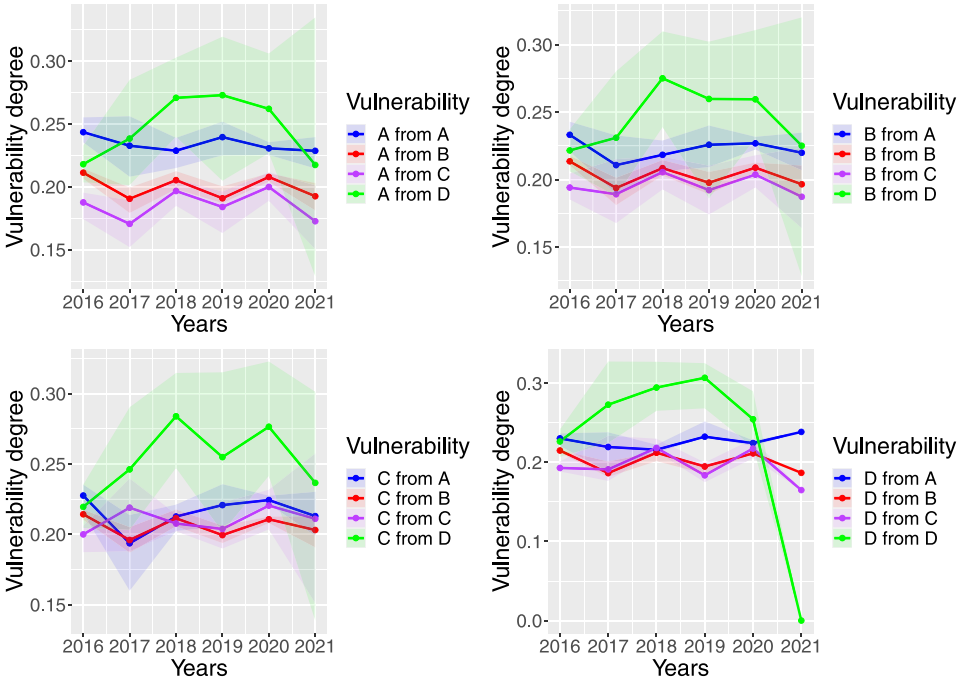


Fig. A.17. Vulnerability degree of the A, B, C and D ESG classes from the A, B, C and D ESG classes, estimated from our dataset which includes $N = 454$ companies and spans the time interval from January 4, 2016 to December 31, 2021.

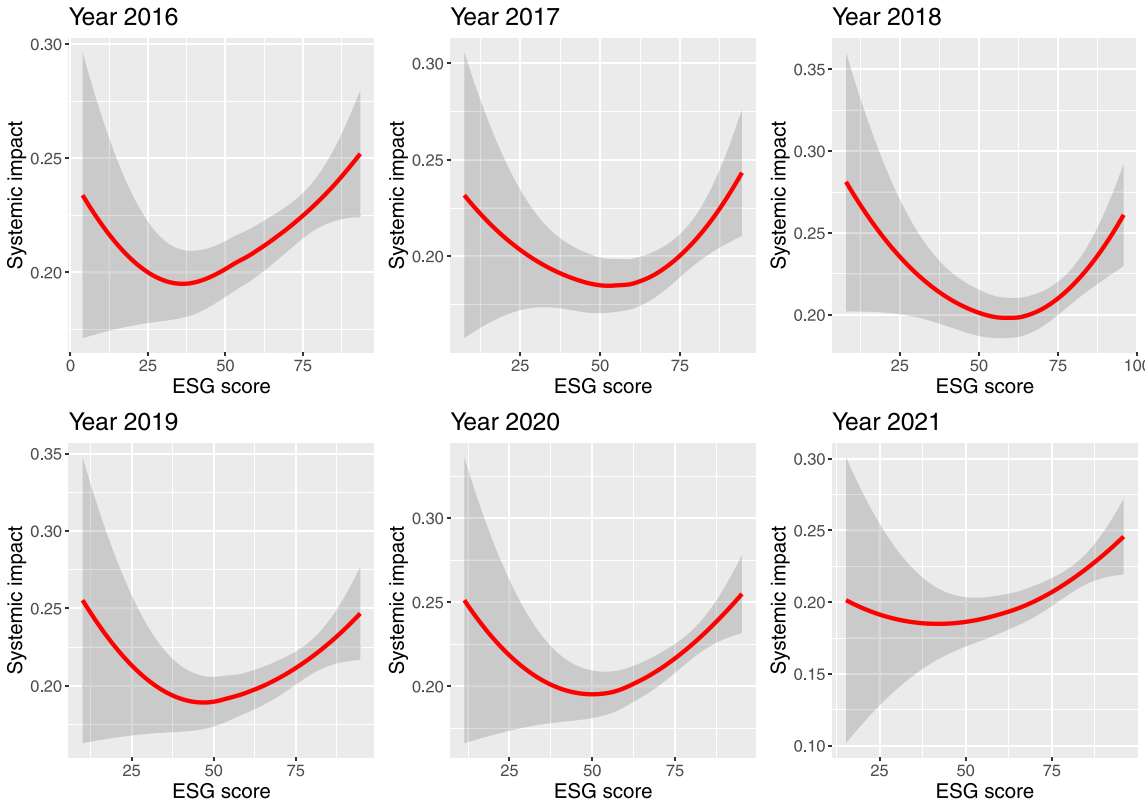


Fig. A.18. Systemic impact defined by the $C^{(10)}_{\bullet \leftarrow j,w}$ indicator as a local function of the ESG score. *Notes*: these estimates are obtained from our dataset which includes $N = 454$ companies and spans the time interval from January 4, 2016 to December 31, 2021. The confidence intervals are delimited by the shaded area.

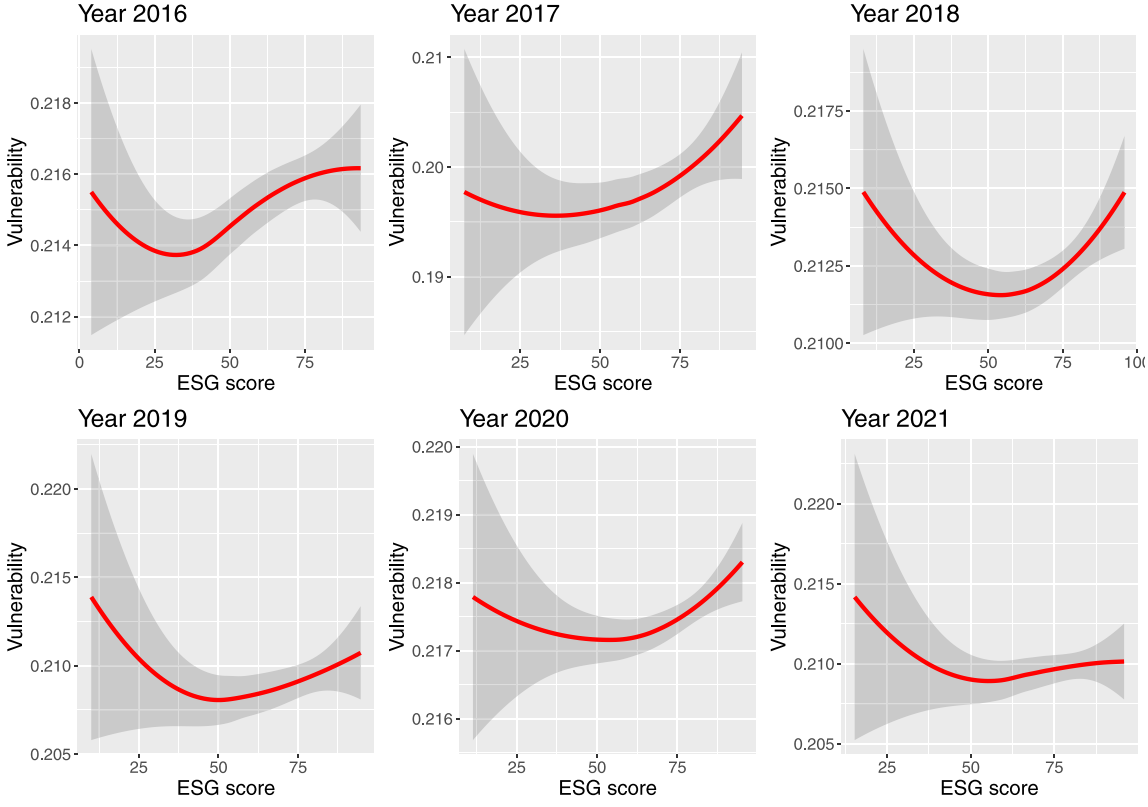
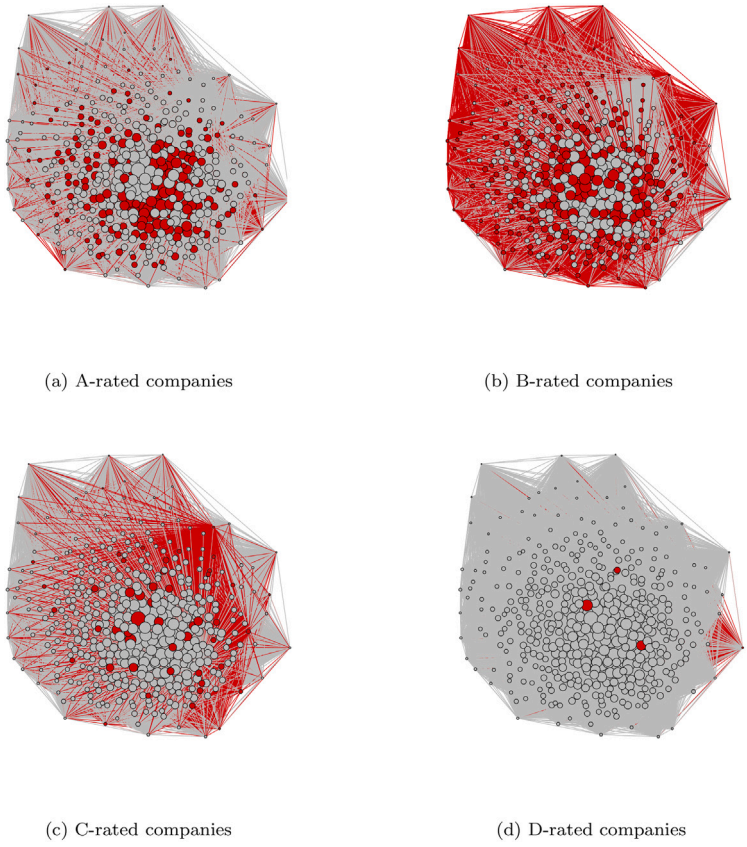


Fig. A.19. Vulnerability defined by the $C^{(10)}_{i \leftarrow \bullet,w}$ indicator as a local function of the ESG score. *Notes*: these estimates are obtained from our dataset which includes $N = 454$ companies and spans the time interval from January 4, 2016 to December 31, 2021. The confidence intervals are delimited by the shaded area.



(a) A-rated companies (b) B-rated companies

(c) C-rated companies (d) D-rated companies

Fig. A.20. The role of the different classes of companies (highlighted in red), clustered by average ESG score, within the network estimated from our dataset which includes $N = 454$ companies and spans the time interval from January 4, 2016 to December 31, 2021. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

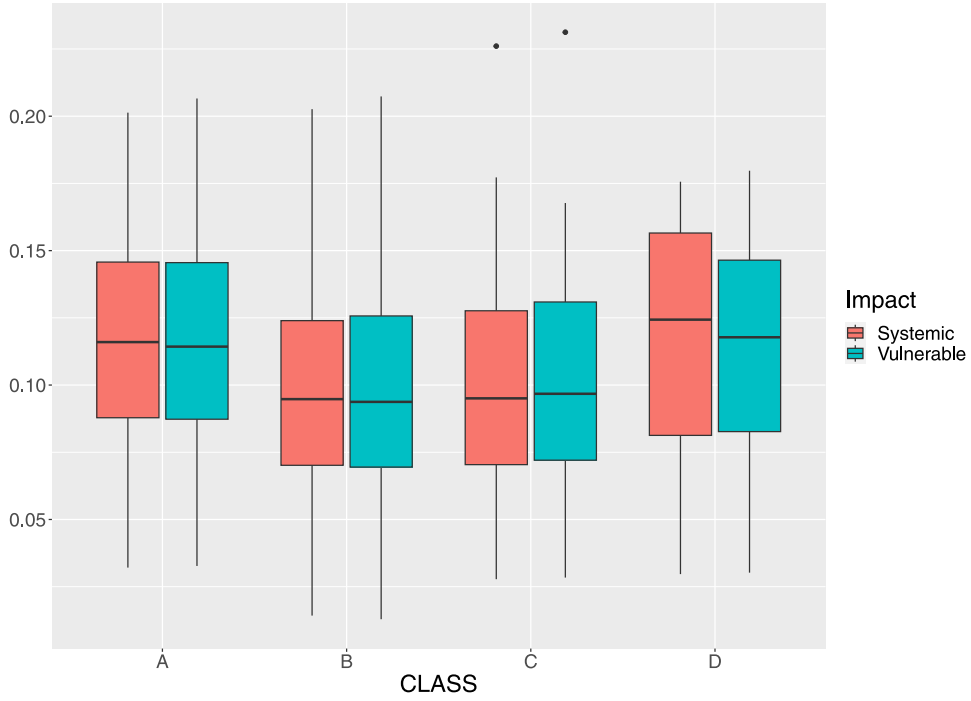


Fig. A.21. Distribution of the $C^{(h)}_{\bullet\leftarrow j}$ (systemic weighted degree) and $C^{(h)}_{i\leftarrow\bullet}$ (vulnerable weighted degree) indicators defined in Eqs. (11) and (10), respectively, clustered by ESG class. *Notes*: these estimates are obtained from our dataset which includes $N = 454$ companies and spans the time interval from January 4, 2016 to December 31, 2021. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

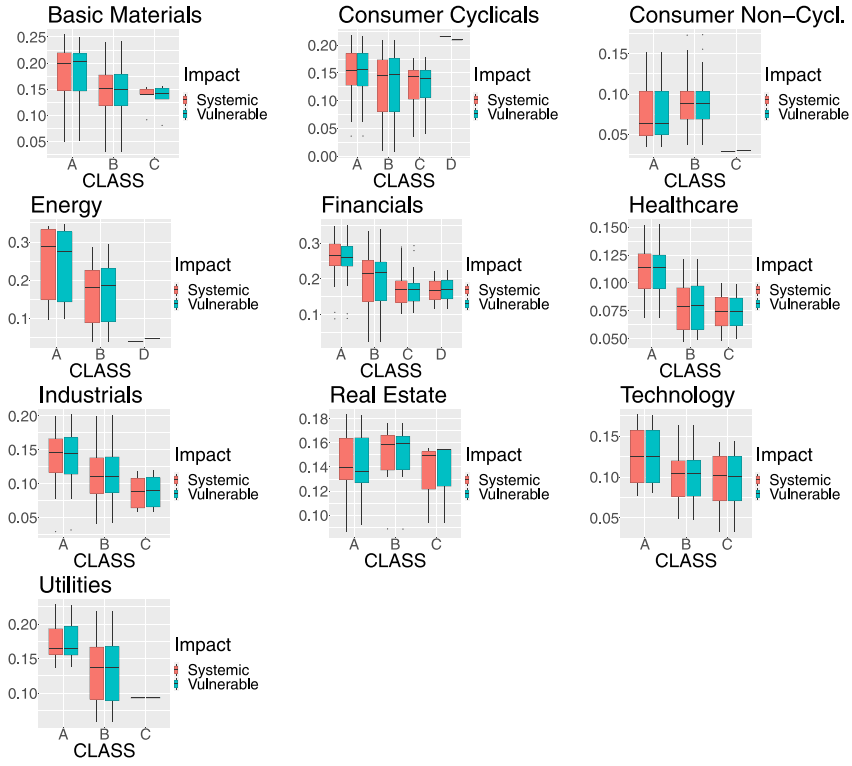
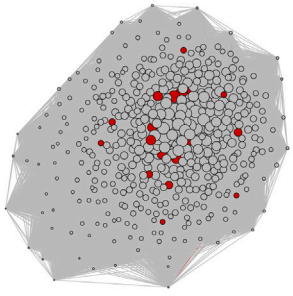
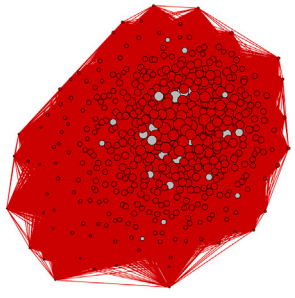


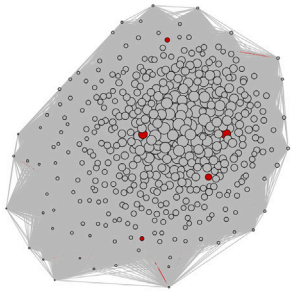
Fig. A.22. Distribution of the $C^{(h)}_{\bullet\leftarrow j}$ (systemic weighted degree, red color) and $C^{(h)}_{i\leftarrow\bullet}$ (vulnerable weighted degree, cyan color) indicators defined in Eqs. (11) and (10), respectively, estimated from the $N = 454$ companies belonging to our dataset clustered by ESG class and economic sector. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)



(a) A—AAA-rated companies



(b) B—BBB-rated companies



(c) CCC-rated companies

Fig. A.23. The role of the different classes of companies (highlighted in red), clustered by average ESG score, within the network estimated from the full-sample data observed from January 4, 2016 to December 31, 2021. *Notes*: these networks are estimated using the return time series and the MSCI ESG ratings provided by Bloomberg. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

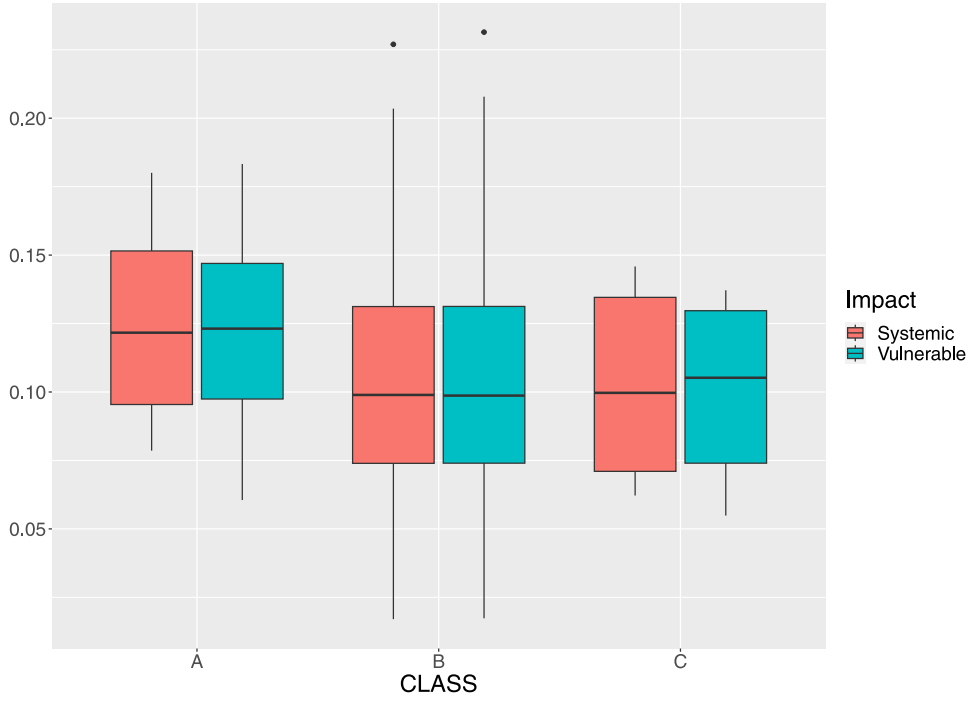


Fig. A.24. Distribution of the $C^{(h)}_{\bullet\leftarrow j}$ (systemic weighted degree) and $C^{(h)}_{i\leftarrow\bullet}$ (vulnerable weighted degree) indicators defined in Eqs. (11) and (10), respectively, clustered by ESG class. *Notes*: these boxplots are obtained using the return time series and the MSCI ESG ratings provided by Bloomberg. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

References

Alessi, L., Ossola, E., Panzica, R., 2021. What greenium matters in the stock market? The role of greenhouse gas emissions and environmental disclosures. J. Financ. Stab. 54, 100869.

Alpaydin, E., 2004. Introduction to Machine Learning. The MIT Press.

Alter, A., Beyer, A., 2014. The dynamics of spillover effects during the European sovereign debt turmoil. J. Bank. Financ. 42, 134–153. http://dx.doi.org/10.1016/j.jbankfin.2014.01.030.

Ameur, H.B., Jawadi, F., Jawadi, N., Cheffou, A.I., 2020. Assessing downside and upside risk spillovers across conventional and socially responsible stock markets. Econ. Model. 88, 200–210.

Andrieş, A.M., Ongena, S., Sprincean, N., Tunaru, R., 2022. Risk spillovers and interconnectedness between systemically important institutions. J. Financ. Stab. 58, 100963. http://dx.doi.org/10.1016/j.jfs.2021.100963.

Apostolakis, G., Papadopoulos, A.P., 2015. Financial stress spillovers across the banking, securities and foreign exchange markets. J. Financ. Stab. 19, 1–21. http://dx.doi.org/10.1016/j.jfs.2015.05.003.

Battiston, S., Dafermos, Y., Monasterolo, I., 2021. Climate risks and financial stability.

Bax, K., Bonaccolto, G., Paterlini, S., 2022. Do lower environmental, social, and governance (ESG) rated companies have higher systemic impact? Empirical evidence from Europe and the United States. Corp. Soc. Responsib. Environ. Manage. 30 (3), 1406–1420. http://dx.doi.org/10.1002/csr.2427.

Belloni, A., Chernozhukov, V., 2011. ℓ1-Penalized quantile regression in high-dimensional sparse models. Ann. Statist. 39 (1), http://dx.doi.org/10.1214/10-aos827.

Berg, E., Lange, K.W., 2020. Enhancing ESG-Risk Modelling-A study of the dependence structure of sustainable investing (Master's thesis). KTH Royal Institute of Technology, URL http://hdl.handle.net/2445/169669, Accessed on 03-11-2021.

Bhattacharya, S., Sharma, D., 2019. Do environment, social and governance performance impact credit ratings: a study from India. Int. J. Ethics Syst. 35 (3), 466–484.

Biau, G., 2012. Analysis of a random forests model. J. Mach. Learn. Res. 13, 1063–1095.

Bonaccolto, G., Borri, N., Consiglio, A., 2023. Breakup and default risks in the great lockdown. J. Bank. Financ. 147, 106308. http://dx.doi.org/10.1016/j.jbankfin.2021.106308.

Bonaccolto, G., Caporin, M., Paterlini, S., 2019. Decomposing and backtesting a flexible specification for CoVaR. J. Bank. Financ. 108, 105659.

Bostanci, G., Yilmaz, K., 2020. How connected is the global sovereign credit risk network? J. Bank. Financ. 113, 105761. http://dx.doi.org/10.1016/j.jbankfin.2020.105761.

Boubaker, S., Cellier, A., Manita, R., Saeed, A., 2020. Does corporate social responsibility reduce financial distress risk? Econ. Model. 91, 835–851.

Breiman, L., 2001. Random forests. Mach. Learn. 45 (1), 5–32. http://dx.doi.org/10.1023/a:1010933404324.

Cerqueti, R., Ciciretti, R., Dalò, A., Nicolosi, M., 2021. ESG investing: A chance to reduce systemic risk. J. Financ. Stab. 54, 100887.

Chen, Y., Lin, B., 2022. Quantifying the extreme spillovers on worldwide ESG leaders' equity. Int. Rev. Financ. Anal. 84, 102425.

Chevallier, J., Nguyen, D.K., Siverskog, J., Uddin, G.S., 2018. Market integration and financial linkages among stock markets in Pacific Basin countries. J. Empir. Financ. 46, 77–92. http://dx.doi.org/10.1016/j.jempfin.2017.12.006.

Chuang, L.-Y., Chang, H.-W., Tu, C.-J., Yang, C.-H., 2008. Improved binary PSO for feature selection using gene expression data. Comput. Biol. Chem. 32 (1), 29–38. http://dx.doi.org/10.1016/j.compbiolchem.2007.09.005.

Demirer, M., Diebold, F.X., Liu, L., Yilmaz, K., 2017. Estimating global bank network connectedness. J. Appl. Econometrics 33 (1), 1–15. http://dx.doi.org/10.1002/jae.2585.

Diebold, F.X., Yılmaz, K., 2014. On the network topology of variance decompositions: Measuring the connectedness of financial firms. J. Econometrics 182 (1), 119–134. http://dx.doi.org/10.1016/j.jeconom.2014.04.012.

Ding, C., Peng, H., 2005. Minimum redundancy feature selection from microarray gene expression data. J. Bioinform. Comput. Biol. 3 (2), 185–205. http://dx.doi.org/10.1142/s0219720005001004.

Eratalay, M.H., Cortés Ángel, A.P., 2022. The impact of ESG ratings on the systemic risk of European blue-chip firms. J. Risk Financ. Manage. 15 (4), 153.

European Commission, 2023. Corporate sustainability reporting. URL https://finance.ec.europa.eu/capital-markets-union-and-financial-markets/company-reporting-and-auditing/company-reporting/corporate-sustainability-reporting_en, Accessed: insert-date-here.

Fan, J., Li, R., 2001. Variable selection via nonconcave penalized likelihood and its oracle properties. J. Amer. Statist. Assoc. 96 (456), 1348–1360. http://dx.doi.org/10.1198/016214501753382273.

Fruchterman, T.M.J., Reingold, E.M., 1991. Graph drawing by force-directed placement. Softw. - Pract. Exp. 21 (11), 1129–1164. http://dx.doi.org/10.1002/spe.4380211102.

Greenwood-Nimmo, M., Nguyen, V.H., Shin, Y., 2023. What is mine is yours: Sovereign risk transmission during the European debt crisis. J. Financ. Stab. 65, 101103. http://dx.doi.org/10.1016/j.jfs.2023.101103.

Gross, C., Siklos, P.L., 2019. Analyzing credit risk transmission to the nonfinancial sector in Europe: A network approach. J. Appl. Econometrics 35 (1), 61–81. http://dx.doi.org/10.1002/jae.2726.

Guyon, I.M., Elisseeff, A., 2003. An introduction to variable and feature selection. J. Mach. Learn. Res. 3, 1157–1182.

Guyon, I., Weston, J., Barnhill, S., Vapnik, V., 2002. Gene selection for cancer classification using support vector machines. Mach. Learn. 46 (1/3), 389–422. http://dx.doi.org/10.1023/a:1012487302797.

Härdle, W.K., Simar, L., 2019. Applied Multivariate Statistical Analysis. Springer International Publishing, http://dx.doi.org/10.1007/978-3-030-26006-4.

Hastie, T., Tibshirani, R., Friedman, J., 2009. The Elements of Statistical Learning. Springer-Verlag GmbH, ISBN: 9780387848587.

Hautsch, N., Schaumburg, J., Schienle, M., 2014. Financial network systemic risk contributions. Rev. Financ. 19 (2), 685–738. http://dx.doi.org/10.1093/rof/rfu010.

Hurvich, C.M., 1985. Data-driven choice of a spectrum estimate: extending the applicability of cross-validation methods. J. Amer. Statist. Assoc. 80 (392), 933–940. http://dx.doi.org/10.1080/01621459.1985.10478207.

Iqbal, N., Naeem, M.A., Suleman, M.T., 2022. Quantifying the asymmetric spillovers in sustainable investments. J. Int. Financ. Mark. Inst. Money 77, 101480.

Iwanicz-Drozdowska, M., Rogowicz, K., Kurowski, Ł., Smaga, P., 2021. Two decades of contagion effect on stock markets: Which events are more contagious? J. Financ. Stab. 55, 100907. http://dx.doi.org/10.1016/j.jfs.2021.100907.

Jay, N.D., Papillon-Cavanagh, S., Olsen, C., El-Hachem, N., Bontempi, G., Haibe-Kains, B., 2013. mRMRe: an r package for parallelized mRMR ensemble feature selection. Bioinformatics 29 (18), 2365–2368. http://dx.doi.org/10.1093/bioinformatics/btt383.

Kanas, A., Molyneux, P., Zervopoulos, P.D., 2023. Systemic risk and CO2 emissions in the US. J. Financ. Stab. 64, 101088.

Kohavi, R., John, G.H., 1997. Wrappers for feature subset selection. Artificial Intelligence 97 (1–2), 273–324. http://dx.doi.org/10.1016/s0004-3702(97)00043-x.

Langley, P., 1994. Selection of relevant features in machine learning. In: AAAI Fall Symposium on Relevance. AAAI Press, New Orleans, pp. 140–144.

Law, M., Figueiredo, M., Jain, A., 2004. Simultaneous feature selection and clustering using mixture models. IEEE Trans. Pattern Anal. Mach. Intell. 26 (9), 1154–1166. http://dx.doi.org/10.1109/tpami.2004.71.

Lazar, C., Taminau, J., Meganck, S., Steenhoff, D., Coletta, A., Molter, C., de Schaetzen, V., Duque, R., Bersini, H., Nowe, A., 2012. A survey on filter techniques for feature selection in gene expression microarray analysis. IEEE/ACM Trans. Comput. Biol. Bioinform. 9 (4), 1106–1119. http://dx.doi.org/10.1109/tcbb.2012.33.

Lütkepohl, H., 2007. New Introduction to Multiple Time Series Analysis. Springer Berlin Heidelberg, ISBN: 3540401725.

Meucci, A., 2005. Risk and Asset Allocation, vol. 1, Springer.

Murè, P., Spallone, M., Mango, F., Marzioni, S., Bittucci, L., 2021. ESG and reputation: The case of sanctioned Italian banks. Corp. Soc. Responsib. Environ. Manage. 28 (1), 265–277.

Murphy, K.P., 2012. Machine Learning. MIT Press Ltd, ISBN: 0262018020.

Pesaran, H.H., Shin, Y., 1998. Generalized impulse response analysis in linear multivariate models. Econom. Lett. 58 (1), 17–29. http://dx.doi.org/10.1016/s0165-1765(97)00214-0.

Pham, S.D., Nguyen, T.T.T., Do, H.X., Vo, X.V., 2023. Portfolio diversification during the COVID-19 pandemic: Do vaccinations matter? J. Financ. Stab. 65, 101118. http://dx.doi.org/10.1016/j.jfs.2023.101118.

Refinitiv, 2023. ESG scores. URL https://www.refinitiv.com/en/sustainable-finance/esg-scores#t-score-range, Accessed: insert-date-here.

Riccobello, R., Bogdan, M., Bonaccolto, G., Kremer, P.J., Paterlini, S., Sobczyk, P., 2022. Sparse graphical modelling via the sorted l_1-norm. http://dx.doi.org/10.48550/ARXIV.2204.10403.

Rothman, A.J., 2012. Positive definite estimators of large covariance matrices. Biometrika 99 (3), 733–740. http://dx.doi.org/10.1093/biomet/ass025.

Tibshirani, R., 1996. Regression shrinkage and selection via the lasso. J. R. Stat. Soc. Ser. B Stat. Methodol. 58 (1), 267–288. http://dx.doi.org/10.1111/j.2517-6161.1996.tb02080.x.

Torri, G., Giacometti, R., Paterlini, S., 2018. Robust and sparse banking network estimation. European J. Oper. Res. 270 (1), 51–65. http://dx.doi.org/10.1016/j.ejor.2018.03.041.

Vapnik, V., 2006. Estimation of Dependences Based on Empirical Data. Springer New York, http://dx.doi.org/10.1007/0-387-34239-7.

Xu, X., Chen, C.Y.-H., Härdle, W.K., 2019. Dynamic credit default swap curves in a network topology. Quant. Finance 19 (10), 1705–1726. http://dx.doi.org/10.1080/14697688.2019.1585560.

Zou, H., Hastie, T., 2005. Regularization and variable selection via the elastic net. J. R. Stat. Soc. Ser. B Stat. Methodol. 67 (2), 301–320. http://dx.doi.org/10.1111/j.1467-9868.2005.00503.x.