Informational Linkages and Stock Return Comovement*

Hriday Karnani†

This draft: December 25, 2024

Abstract

We empirically study how excess comovement in stock returns can be explained by stock-to-stock interdependence through informational links. Informational linkages emerge from anomalous interdependence in agents' beliefs about stocks' economic performance. We propose a novel measure for informational linkages based on agents' learning process, which is biased towards common *information* due to learning frictions, generating correlated beliefs across stocks. We empirically measure these correlated beliefs based on analysts' forecast errors. Our results show that informational connections explain stock returns after cleaning for fundamental connections and controlling for various explanations already studied in the literature. We use our estimated informational linkages to study the propagation of a climate event and simulated shocks through the stock market, finding quantitatively important indirect effects.

JEL Codes: D80, G12, G14.

Keywords: Informational interdependence, learning process, comovement, asset pricing.

*I am grateful to Alejandro Bernales and Paula Margaretic for their guidance, time, support, and patience in this process. I also thank Eugenia Andreasen, Jorge Miranda-Pinto and Olga Kolokova for very interesting discussions, and Diego Zuñiga for his help in the initial stage of this project. I acknowledge financial support from Fondecyt Project #1231660. All remaining errors are mine.

†Facultad de Economía y Negocios, Universidad de Chile. Email: hkarnani@fen.uchile.cl

1. Introduction

Stock returns exhibit high correlations even after controlling for fundamental connections (Pindyck and Rotemberg, 1993; David and Simonovska, 2016). This 'excess comovement' suggests that non-fundamental connections play a significant role in explaining stock returns. Among these factors, agents' learning process about asset-related information, the formation of beliefs, and the diffusion of market information have arisen as important elements in understanding stock return dynamics beyond fundamental movements.

This paper empirically explores how learning frictions in agents' learning process can lead to correlated beliefs across stocks, ultimately contributing to excess comovement in the stock market. We do this by using a novel measure for informational linkages based on analysts' forecast errors. Informational linkages reflect anomalous interdependence in agents' beliefs about stocks' economic performance, even if those stocks' are not fundamentally connected. Using this empirical measure, we construct stock-to-stock *informational* linkages and examine whether these linkages show signs of interdependence between stocks, and excess comovement at an aggregate level.

The intuition behind the empirical measure we use is based on the stylized model in Bernales et al. (2024), where agents receive new information about the unknown future values of economic indicators across different units (e.g., countries, stocks). Due to learning frictions (such as learning costs and limited attention capacity) agents overweight new information about factors that simultaneously impact multiple units. This overweighting leads to informational linkages, which can be seen as excessive correlations in agents' forecast errors regarding the economic variables that characterize different units. Without learning frictions, forecast errors for economic variables of different stocks should not be correlated after controlling for fundamental connections, and informational linkages would not exist.

To quantify these informational linkages, we calculate analysts' average one-year ahead Earnings Per Share (EPS) forecast errors following a diverse set of US stocks, at a monthly frequency from 2000 to 2019. We then clean these forecast errors from factors that could affect forecasts for all stocks (such as economy-wide shocks). Subsequently, we compute the correlations of the cleaned average one-year ahead EPS forecast errors between pairs of stocks, keeping only the most relevant connections. This process is the basis of our stock-to-stock informational linkage matrix.

We are interested in studying if the proposed informational channel can explain the component of stock returns that fundamental associations do not explain, that is, idiosyncratic returns. We use a multi-factor Fama-French model to estimate idiosyncratic returns. In a standard asset pricing framework, idiosyncratic returns should be uncorrelated and independent across stocks.

We test the relevance of informational linkages in explaining interdependence between stocks, following a Spatial Two-Stages Least Squares (S2SLS) methodology. We show that informational linkages explain idiosyncratic stock returns, providing evidence of interdependence between stock returns.

We address endogeneity concerns in several ways. First, to mitigate endogeneity concerns between the construction of informational linkages and stock returns, we use one-year ahead forecast errors on EPS instead of on stock returns. Second, we we allow the informational linkages interdependence matrix to vary every year and use the lagged matrix of interdependence in each period, to avoid simultaneity between the construction of the informational linkages matrix and stock returns.

Third, we estimate our model using an instrumental variable procedure, a Spatial Two-Stage Least Squares (S2SLS) methodology, the intuition behind it is that we instrument the stock returns that are informationally linked to a stock i with fundamental factors of those stocks linked to i. Fourth, we aim to avoid any reverse causality issues in our methodology by including several lagged control variables typically used in the finance literature to explain stock returns. Fifth, we address the potential of cross-sectional dependence in forecast errors by cleaning (de-factoring) analysts' forecast errors from factors that hit the whole stock market, this also accounts for any *unobserved* global fundamental factors.

We perform additional validation results by studying how our measure of informational interdependence behaves across various scenarios, based on work done in previous studies (Barberis and Shleifer, 2003; Barberis et al., 2005; Calvo and Mendoza, 2000; Israelsen, 2016; Peng and Xiong, 2006; Veldkamp, 2006), we study how learning frictions (such as the costs related to learning about different stocks) can induce agents to learn in terms of categories or groups of stocks instead of following stock-specific information, which should also generate informational linkages. We find that our measure of informational linkages is important in explaining stock returns even after considering potential categories in which agents could group stocks.

The production of information across time is important in the learning process of agents. In stressful periods, agents have more at stake, which should increase the production of information. This will impact the learning frictions that agents face, specifically, the costs of learning about each specific stock, and thus, the formation of informational linkages. We find that informational linkages behave differently in periods with higher uncertainty. In particular, the intensity of the informational channel decreases, consistent with the intuitions behind Gorton and Ordoñez (2014).

We acknowledge the potential of common analysts' influence on our proposed measure of informational interdependence. The informational connections captured by our measure may be affected by the presence of common analysts providing EPS forecasts for multiple stocks. Analysts' forecasts can be biased due to factors such as career concerns or varying amounts of information available for different stocks. However, we address this concern through several aspects of our methodology.

First, since we use forecast *errors* rather than the predicted EPS values, biases in forecast errors issues should not be a big concern. Second, our 'de-factoring' process for analysts' forecast errors implicitly cleans for analyst consensus bias induced by career concerns. Third, by using correlations of forecast errors to measure informational links, we further diminish potential biases, since the 'level' of the forecast error is not relevant anymore. Fourth, if the amount of information available for two stocks differs significantly, it should reduce the correlation between their forecast errors, naturally accounting for information asymmetries.

We then move into applications where we analyze the impacts of shocks in the stock market through our estimated informational linkages. This is important because according to our model, a shock to a stock i will not only affect stock i, but also those stocks that are directly and indirectly connected to the stock i through our matrix of information links. To do so, we estimate the propagation (indirect effects) through the stock market of simulated and real-world shocks to one (or a small amount) of stocks. First, we simulate a shock to one stock and analyze how it propagates to the stock market through our informational linkages. We find evidence that the indirect effects of these shocks, the effect of a shock over stocks not directly affected by it, are quantitatively important.

Continuing with the analysis of shock propagation, we also study how heterogeneous the previous result is depending on how central each stock is. The more central firms are those that are more connected[1]. Central stocks should propagate shocks more than others; we quantify this effect.

Finally, we also analyze if our informational linkages matrix is able to capture how real-world shocks propagate. We take a climate event that *directly* affected few states. First, we show that stocks in those states had negative stock returns, while the rest of the stock market didn't. Then, we show that stocks connected through our informational linkages matrix to stocks that suffered from the climate event also suffered from quantitatively important negative stock returns. This provides evidence that our informational linkages matrix is able to capture the indirect effects of real-world shocks.

The rest of the paper is organized as follows. Section 2 reviews what has been studied in the literature and how we contribute to it. Section 3 explains the intuitions behind the informational channel and the measure we use to quantify them. Section 4 explains in detail our empirical strategy and the data used. Section 5 discusses the results and the implications that come out of them. Section 6 studies the simulation and real-world propagation of shocks through our estimated network, and Section 7 concludes.

2. Literature Review

Previous literature provides evidence suggesting that even professional forecasters do not fully adhere to the traditional full-information hypothesis of rational expectations (Coibion

[1]Connections are not necessarily only of first order. If the stock "A" is connected to the stock "B" and the stock "B" has many connections, then the stock "A" will also be indirectly connected to those connections of the stock "B".

and Gorodnichenko, 2015). Building on this idea, Coibion et al. (2018) and Coibion et al. (2019) also show that firms' and professional forecasters' beliefs affect their economic decisions[2]. We build on this literature by showing that friction in agents' learning process can induce correlated beliefs between stocks.

We also add to the literature studying frictions in the learning process of agents. Veldkamp (2006) develops a framework where agents need to learn about assets in the economy, but they have limited attention capacity. It is optimal for them to consume a subset of information in the economy, inducing stock market comovement. Learning frictions can also induce category-learning behavior, where agents learn in terms of what they think are similar groups, as described in Barberis and Shleifer (2003); Peng (2005); Peng and Xiong (2006). These studies provide a theoretical foundation on why observing excess comovement can be an optimal equilibrium result, but they are typically focused on a couple of assets, whereas we focus on testing this hypothesis empirically for most of the assets in the economy.

Our definition of informational linkages, the theory and mechanisms we postulate behind them is heavily based on Bernales et al. (2024), which studies how informational linkages can transmit shocks between countries, even after controlling for real and financial linkages. We instead focus on the effect of informational linkages in explaining interdependence between stocks, at a micro-level[3].

We also contribute to the extensive literature studying stock market comovement, particu-

[2]Giglio et al. (2021) also provides new evidence on the link between beliefs and real actions taken by survey respondents to show that beliefs are shown in portfolio allocations. Conditional on trading, beliefs affect both the direction and the magnitude of trading.

[3]There are few other papers that study the effects of informational linkages on the economic transmission of shocks, such as Kaminsky and Schmukler (1999); Basu (2002); Colla and Mele (2010). Opposed to us, the basis of their measures of informational linkages is in news instead of agents' beliefs.

larly showing evidence that comovement in the stock market is not completely driven by fundamental connections or rational behavior (Lee et al., 1991; Pindyck and Rotemberg, 1993; Barberis and Shleifer, 2003; Barberis et al., 2005; Veldkamp, 2006; Greenwood, 2007; Kumar et al., 2016; David and Simonovska, 2016), we do this by showing evidence of interdependence and excess comovement in the stock market that can be determined by anomalous comovements in agents' belefs.

There has been an important amount of studies proposing mechanisms that explain the excess comovement puzzle in the literature (Barberis et al., 2005; Antón and Polk, 2014; Muslu et al., 2014; Hameed et al., 2015; Israelsen, 2016). These mechanisms vary from category learning views of investors, common analyst coverage that induces correlated beliefs, mutual fund ownership, to many others. We show that many of these theories complement the informational channel that we propose in explaining stock return excess comovement.

We also contribute to the literature studying the dynamics of optimal production and acquisition of information, which can also shed light on how agents' behavior changes when there is 'more at stake' (such as crisis periods), compared to normal periods (Gorton and Ordoñez, 2014; Chousakos et al., 2023; Cole et al., 2024)[4]. We contribute to this literature by providing empirical evidence that the intensity of our informational linkages (based on the correlation of beliefs between stocks) diminishes in periods with higher uncertainty, thus, reducing learning biases in periods where agents have more at stake.

Finally, our study can also be related to the literature studying herding behavior and contagion in markets, as in Scharfstein and Stein (1990); Froot et al. (1992); Calvo and

[4]The idea behind this is that producing information is costly and it is not optimal to produce information at every period. When there are negative shocks, information is more valuable since it allows us to learn the true quality of the investments being held.

Mendoza (2000); Bekaert et al. (2014); Debarsy et al. (2018). Herding can be viewed as the aggregate result of the interdependence between groups of stocks, which we find. On the other hand, with our methodology, we can also study the contagion of simulated and real shocks in the stock market.

3. The Informational Channel

This section presents the intuitions and the empirical measure used to quantify our informational channel approach. We begin by explaining the intuitions that motivate our framework, based on the literature on learning frictions in agents' learning process. We then detail our method for quantifying informational linkages between stocks, which is based on analysts' forecast errors. This approach allows us to construct a network of stock-to-stock informational connections that capture the interdependence arising from correlated beliefs among agents.

3.1. Intuitions behind the Informational Channel

Our motivating framework is based on the literature describing learning frictions in agents' learning process (Peng and Xiong, 2006; Veldkamp, 2006; Bernales et al., 2024). We hypothesize that learning frictions in agents' learning process, such as limited attention capacity and costs to learn new information, induce correlated beliefs about the variables that describe the economic performance of stocks. We name the formation of these correlated beliefs informational linkages.

These links capture many ways in which stocks can be linked with each other, some of them can not be directly observed by the data, which is one of the novelties of our approach.

Directly, these links capture common beliefs among agents. We rely on the stylized model proposed by Bernales et al. (2024), which is easily applicable to this context, to explain the intuitions and mechanisms through which the informational channel arises.

Adapting their stylized model to our context, a representative agent wants to learn about the unknown values of a variable that characterizes the performance of a stock (for instance, the return of a stock). Stock returns are determined by a linear combination of two factors: a stock-specific factor (idiosyncratic) and a common factor between stocks. These factors are identically and independently distributed (i.i.d.) and follow normal distributions. Each factor has a specific mean and variance, unknown by the agent. Agents receive signals (information) about each factor. Each of these signals has a specific precision, which is unknown but inferred by the agent.

The agent learning process is determined by how they interpret the signal regarding each factor. They assume that the agent over-weights the precision of the common factor signal. This over-weighting arises due to learning frictions in the agents learning process, such as (i) *learning costs*, since the cost of learning about each stock is usually higher than the cost of learning about a common set of the market (Veldkamp, 2006), and (ii) *limited attention capacity*, which induces them to learn in terms of categories/groups (Barberis et al., 2005; Peng and Xiong, 2006).

Motivated by this intuition, agents wrongly believe that the signal regarding the common factor is more informative than it is in reality, over-weighting the precision of this signal. This bias in the learning process will generate the agent to have correlated beliefs across stocks, which will create interdependence in idiosyncratic stock returns, and comovement in the stock market. The formation of these correlated beliefs is what we call the informational channel.

$$\text{new info.} \rightarrow \underbrace{\text{learning process}}_{\text{biased towards common factors}} \rightarrow \underbrace{\text{correlated errors in agents' beliefs}}_{\text{informational linkages}} \rightarrow \text{comov. stock returns}$$

3.2. Quantifying Informational Linkages

Now we turn into discussing the empirical metric used to construct stock-to-stock informational linkages. The intuition behind this metric is that a pair of stocks is informationally connected if the correlation between agents' beliefs is sufficiently strong in a statistical sense. We base this measure of informational links on Bernales et al. (2024), which comes from the stylized model they propose.

To proxy agents' beliefs, we use monthly analyst forecast errors on the Earnings Per Share (EPS)[5] of each stock. Specifically, we use the Normalized Mean Root Squared Error (NMRSE) of the one-year ahead monthly EPS forecast of analysts by stock, estimated as seen in Equation (1).

$$NMRSE_{i,t}^{EPS} = \sqrt{\frac{1}{K}\sum_{k=1}^{K}\left(\frac{E\hat{P}S_{k,i,t+1} - EPS_{i,t+1}}{EPS_{i,t+1}}\right)^{2}} \tag{1}$$

where k is the analyst index, $E\hat{P}S_{k,i,t+1}$ is k's analyst forecast for stock i at period $t+1$, $EPS_{i,t+1}$ is the realized value[6].

The forecast errors observed in the previous step could be driven by common factors (cross-sectional dependence), such as a common shock to the economy or a specific group of it, that made forecasts mistaken for all stocks. If forecast errors are driven by cross-sectional

[5] We use forecast over Earnings Per Share and not stock returns to avoid endogeneity issues when estimating interdependence between stock returns.

[6] To avoid results being driven by a small number of analysts, we require a minimum of five analysts to calculate $NRMSE_{i,t}^{EPS}$.

dependence, then our informational linkages measure is not going to be a clean measure of stock-to-stock interdependence, but it would also include common factors from the stock market.

We follow the two-stage methodology proposed by Bailey et al. (2016) to correct for cross-sectional dependence by de-factoring the forecast error. The first step is to test for cross-sectional dependence (CSD). If the degree of CSD is high (more than 0.5), we de-factor the measure until this is weakly CSD (below 0.5)[7]. In our case, this is achieved by de-factoring with one common factor, by estimating the following specification:

$$NMRSE_{i,t}^{EPS} = \alpha_i + \lambda_i \mathbf{f}_t + \varepsilon_{i,t}^{DFact} \tag{2}$$

$\mathbf{f}_t$ are unobserved common factors and λ_i are factor loadings[8], which are estimated by principal component analysis (PCA). The defactored observations, $\varepsilon_{i,t}^{DFact}$, are kept to construct our informational linkage matrix.

Then, we estimate the correlations of beliefs between securities and keep those relevant connections. We define $\hat{\epsilon}_{i,j}$ as the sample estimate of the de-factored pairwise correlation of the one-year ahead EPS forecast error between any two securities i and j over $t \in \{1, 2, \ldots, T\}$. In order to keep relevant connections, we follow Bailey et al. (2016) and apply a multiple testing procedure (Holm, 1979):

$$w_{i,j} = I\left(\hat{\epsilon}_{i,j} > T^{-1/2} c_p(n_{ij})\right), \tag{3}$$

[7] Bailey et al. (2016) shows that if the degree of CSD is below 0.5, and the number of observations is sufficiently large (as in our case), the bias that is generated due to the presence of common factors tends to 0.

[8] Our baseline model uses one common factor. However, we have tested variations of this number

where $c_p(n_{ij}) = \Phi^{-1}\left(1 - \frac{p}{2f(n_{ij})}\right)$, p is the size of the test, which is set to $p = 0.01$, $\Phi^{-1}(.)$ is the inverse cumulative standard normal distribution and $f(n_{ij})$ is a function linearly increasing in n_{ij}, the number of firm pairs, that is $n_{ij} = N \times (N-1)$.

The intuition behind this is that an element $w_{i,j}$ of the matrix of informational linkages $\left(\mathbf{W}^I\right)$ is equal to one when the informational link between stocks i and j is sufficiently strong, that is, when it is statistically different from zero. We allow for different degrees of intensity of the link based on $\hat{\epsilon}_{i,j}$, multiplying $w_{i,j}$ by $\hat{\epsilon}_{i,j}$ and row-normalizing $\left(w^I_{i,j} = \frac{\hat{\epsilon}_{i,j} w_{ij}}{\sum_j \hat{\epsilon}_{i,j}}\right)$. Note that this procedure will, by construction, give us a highly sparse (several zero elements) matrix. We want this to happen since we do not want to force informational connections between stocks that are not sufficiently strong.

This procedure gives origin to a $N \times N$ matrix of informational linkages that will have the following structure:

$$W^I = \begin{bmatrix} 0 & w^I_{1,2} & w^I_{1,3} & \dots & w^I_{1,N} \\ w^I_{2,1} & 0 & w^I_{2,3} & \dots & w^I_{2,N} \\ w^I_{3,1} & w^I_{3,2} & 0 & \dots & w^I_{3,N} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ w^I_{N,1} & w^I_{N,2} & w^I_{N,3} & \dots & 0 \end{bmatrix}$$

Note that each element w^I_{ij} is zero when $i = j$, this means that a stock is not informationally connected to itself. Elements $i \neq j$ will be different to zero when $\left(\hat{\epsilon}_{i,j} > T^{-1/2} c_p(n_{ij})\right)$ (i.e., if the informational connection between $i \neq j$ is sufficiently strong in a statistical sense).

4. Empirical Strategy and Data

This section details the empirical strategy used to test the relevance of the informational channel introduced in the previous section. In Section 4.1, we present our approach to study the impact of informational linkages on stock return excess comovement, using spatial econometric models. These models are particularly well-suited for our analysis due to their ability to accommodate the high-dimensional nature of our informational linkage matrix and explicitly model interactions between stocks. This approach also allows us to include a single or multiple sources of interdependences with one or several interaction matrices (i.e., control for other interdependences besides the informational channel proposed), it enables us to analyze the spillover mechanisms from multiple sources of transmission of shocks across securities in a single model, among many other interesting features explained with more detail by LeSage and Pace (2009). Finally, we detail the data used in this analysis in Section 4.2.

4.1. Empirical Strategy

Having introduced the intuition and the metric we use to measure informational linkages between stocks, we move into describing how we empirically estimate the relevance of the estimated informational linkages in explaining stock return excess comovement.

A part of stock market comovement can be explained by fundamental connections. We try to understand the component of stock returns that these fundamental connections should not explain, the idiosyncratic component of stock returns. We follow traditional finance theory and extract the idiosyncratic component of stock returns by estimating a

6-factor Fama-French Model (Fama and French, 1993, 2018) and keeping its residuals. See Appendix A to check the steps we follow to extract the idiosyncratic component of stock returns, which we denote by $\mathbf{r}_t$.

Standard asset pricing theory predicts that the idiosyncratic component of returns between stocks should not be correlated. Despite this prediction, we show in the following sections that the idiosyncratic component of stock returns is correlated, and those correlations can be explained by our proposed informational linkages.

Consider N firms over T periods. Denote by $\mathbf{r}_t$ the vector of idiosyncratic returns in period $t \in T$ for a set of stocks (indexed by i). The model to be estimated is[9]:

$$\mathbf{r}_t = \rho \mathbf{W}^I_{t-1}\mathbf{r}_t + \mathbf{X}_{t-1}\beta_i + \boldsymbol{\mu} + \gamma_t + \mathbf{v}_t \tag{4}$$

Where $\boldsymbol{\mu}$ and γ_t are stock and time fixed effects, $\mathbf{W}^I_{t-1}$ is our $N \times N$ matrix of informational linkages (previously estimated) that varies each year[10], $\mathbf{X}_{t-1}$ is a matrix that contains financial ratios and lagged returns and risk as controls. Note that β_i is security-specific. Thus, we estimate the specific impact of controls over each security. $\mathbf{v}_t$ is an error term identically and independently distributed with mean 0 and variance σ_i^2. It is important to highlight that the diagonal elements of $\mathbf{W}^I_{t-1}$ are zero. Thus, the idiosyncratic return of a

[9]Throughout the paper we follow the matrix notation, but it can also be re-written in a scalar notation:

$$r_{i,t} = \rho \sum_{j=1}^{N} w^I_{ij,t-1} r_{j,t} + \beta_i \sum_{k=1}^{K} x_{i,k} + \mu_i + \gamma_t + v_{it}$$

where k is one of the K regressors in $\mathbf{X}_{t-1}$.

[10]Each value in the informational linkage matrix indicates if a pair of stocks is linked. If they are, these two securities are interconnected. If a security i suffers a shock, it will transmit that shock to securities directly and indirectly connected to i. These interconnections give origin to some unexplained idiosyncratic stock comovement, which motivates this paper.

stock i is not used as an explanatory variable for the same stock i.

We want to understand the variation in idiosyncratic asset returns that can be explained by informational linkages, which is captured by ρ, ρ is a measure of the intensity of informational linkages in explaining idiosyncratic stock returns. If ρ is statistically different from zero, we have evidence of the interdependence/shock transmission across securities.

One concern that can arise is that correlated beliefs about a pair of stocks could arise simultaneously with what determines idiosyncratic returns. Note that $\mathbf{W}^I_{t-1}$ is lagged one year with respect to $\mathbf{r}_t$, we do this to address this concern. We also construct $\mathbf{W}^I_{t-1}$ using monthly data of correlated beliefs across stocks for five years. Thus, we try to capture constructed informational linkages that are persistent, not driven just by one period, but also dynamic. As a robustness check, we also test with a version of $\mathbf{W}^I$ that is fixed across time.

Idiosyncratic stock returns can also be explained by the firm's fundamentals, we use firm-specific control variables for this potential explanation in $\mathbf{X}_{t-1}$. We follow Leary and Roberts (2014) and include Log(Sales), Cash-Assets, Ebitda-Assets, Debt-Assets, and Market-Book, all lagged to the previous period to avoid endogeneity issues due to simultaneity. We also include lagged return and risk as controls, following standard practices in the literature.

We estimate the spatial model proposed following a Spatial Two-Stage Least Squares (S2SLS) methodology following Kelejian and Prucha (1998, 1999, 2004). This is an instrumental variable approach that accounts for endogeneity in the dependent variable[11]. This

[11] $\mathbf{r}_t$ is our dependent variable (LHS of Equation (4), but note that in our specification, the idiosyncratic returns of stocks $i \neq j$ are used (weighted) to explain the idiosyncratic return of stock i. ($W^I_{t-1}\mathbf{r}_t$ the RHS of Equation (4)), since we want to understand if there is interdependence between stock returns. The endogeneity that we are addressing in our S2SLS methodology comes from using $W^I_{t-1}\mathbf{r}_t$ as a regressor.

could raise some endogeneity concerns. The intuition behind the S2SLS methodology is that the idiosyncratic return of a stock i is instrumented with the fundamental characteristics (given in X_{t-1}) of the stocks linked to stock i. In Appendix B, you can find details of this estimation method. The first stage is:

$$W^{I}_{t-1}\mathbf{r}_t = \theta \mathbf{H_n} + \mathbf{u}_t \tag{5}$$

Where $\mathbf{H_n} = \left[\left\{ \left(\mathbf{W}^{I}_{t-1} \otimes \mathbf{I_T} \right) \mathbf{X}_{t-1} \right\}_{t \in \mathbf{T}}, \mathbf{X}_{t-1} \right]$ and the estimated parameters are $\widehat{\theta} = \left[\begin{array}{cc} \widehat{\rho} & \widehat{\beta}_i \end{array} \right]'$. Then, we use $\widehat{W^{I}_{t-1}\mathbf{r}_t}$ in the second stage as in Equation (4).

Although we have explained why we chose Spatial Econometric Models to estimate empirically our hypothesis, it is valid to question it. To clear some concerns, we compare this methodology to its potential competitor, Global-VARs (GVAR).

As described by Elhorst et al. (2021), the GVAR literature has traditionally focused mainly on interactions among countries, in the particular context where $T > N$. Our case is different; we have $N > T$. Moreover, in GVARs, there is usually no distinction between different types of variables in the sense that they are all treated as dependent variables in a system of equations. We do control for stock fundamentals as independent variables and we do not want them to be treated as dependent.

Under some circumstances, a spatial econometric model is shown to be a special case of a GVAR model as it tends to be univariate rather than multivariate and has homogeneous rather than heterogeneous coefficients. This is of particular interest for us, since we would like to capture how relevant informational linkages as an aggregate are in explaining stock market excess comovement.

Also, GVAR models tend to treat the spatially lagged dependent variable ($W\mathbf{r}$ in our case) as exogenous if W is sufficiently dense. Spatial econometric models focus on potentially sparse levels of connections, as opposed to GVARs. Elhorst et al. (2021) suggests that when one has a model with the characteristics that we have pointed out in this section, it is better to use Spatial Econometric Models rather than GVARs.

4.2. Data

In terms of data, we use forecasts from analysts on the Earnings Per Share of different stocks from I/B/E/S, financial ratios of firms from Compustat, and security returns from CRSP. We estimate the spatial model with data from January 2005 until December 2019. However, to construct each $\mathbf{W}^I_{t-1}$ with $t \in \{2005, ..., 2019\}$, we calculate the correlations of the residualized analyst forecast errors of the previous 5 years. Thus, $\mathbf{W}^I_{2005}$ uses data from 2000 until 2004, $\mathbf{W}^I_{2006}$ uses data from 2001 until 2005, and so on. This means that we require data from I/B/E/S from 2000 until 2018 for each security[12].

We keep as many stocks as possible. Due to our methodology, we require securities to be present in most of the timespan of our estimation. Considering these restrictions, we keep a subset of 771 securities. About half of them have been part of the Standard & Poor's Index at some point, and half of them have never been part of it.

Table 1 shows a summary statistics of our subset of securities[13].

[12]This requirement is the main one that restricts us from increasing the number of stocks used in our analysis. We tried to use as many stocks as possible, but many stocks don't live for so many years. Concerns about selection in our sample of stocks can arise due to this. We acknowledge that this assumption can have an impact on our results, but it is not clear if the impact is positive or negative, and it is the best we can do with the data we have.

[13]Note that it is okay if the number of observations of forecast errors is lower than the others. We extract correlations between those forecast errors, we do not need this variable to be part of a balanced panel.

Table 1. Summary Statistics

	Min.	p25	p50	p75	Max.	Mean	St.Dev.	Obs.
Idios. Return	-0.18	-0.04	-0.00	0.04	0.20	-0.00	0.09	136,980
Log(Sales)	11.28	13.72	14.81	15.93	18.37	15	1.7	136,807
Cash-Assets	0.00	0.02	0.07	0.18	1.02	0.25	2.2	136,721
Ebitda-Assets	-0.41	0.27	0.72	1.74	149.36	23	332	136,704
Debt-Assets	0.12	0.42	0.58	0.74	0.93	0.57	0.22	136,819
Market-Book	0.62	1.36	2.16	3.52	12.99	3.6	15	136,712
EPS Forecast Errors $(NMRSE_{i,t-1})$	0.02	0.07	0.16	0.39	6.18	0.52	1.1	118,954

Notes: Idios. Returns is the residualized monthly returns of stocks after applying a Fama-French 6 Factor Model to excess returns (return-risk free rate). Analys. EPS Forecasts Errors are estimated following Eq. 1. Every variable is winsorized at the 2% of each tail to avoid estimations being driven by outliers.

5. Results

In this section, we check the relevance of the informational channel. Specifically, we check if the constructed matrix of informational linkages can explain, in a statistical way, idiosyncratic returns of securities. If it does, there is evidence of interdependence between securities, which would shed light on how stock market excess comovement arises. In Section 5.1 we analyze the relevance of the informational channel. In Section 5.2 we perform some validity analysis by controlling for other potential channels studied in the literature. In Section 5.3 we analyze the robustness of our results.

5.1. Relevance of the Informational Channel

Table 2 shows the relevance of the informational channel. The coefficient shown is ρ from the baseline specification. As it can be seen in the table, ρ is positive and statistically significant. This table tells us that a shock in a security i does not only impact security i but also all securities $i \neq j$ linked directly and indirectly in an informational way to security i.

This is strong evidence that the informational channel can explain idiosyncratic stock returns and stock market excess comovement. It is relevant to mention that the interpretation of coefficients in spatial models is not a typical marginal effect[14]. Intuitively, the coefficient ρ is a measure of the average strength of informational spillovers across stocks (LeSage and Thomas-Agnan, 2015).

To be sure that results are not driven by the control variables selected, we estimate our baseline model described in Equation (4) with different sets of controls. The table shows that the estimated coefficient do not vary depending on the set of controls used.

It is important to mention that we find strong evidence that the informational channel proposed can explain interdependence between idiosyncratic stock returns even after taking into account various potential bias concerns in our results.

To mitigate endogeneity issues between informational linkages and stock returns, we utilize one-year ahead forecast errors on EPS rather than stock returns themselves. Additionally, we allow the informational linkages interdependence matrix to vary annually,

[14] Note that Equation (4) can be also written as:

$$\mathbf{r}_t = (I_N - \rho \mathbf{W}^I_{t-1})^{-1}(\mathbf{X}_{t-1}\beta_i + \boldsymbol{\mu} + \gamma_t + \mathbf{v}_t)$$

Thus, the impact of a change on the characteristics (say X) of a stock on the outcome ($\mathbf{r}_t$) is not only β, the indirect effects over other stocks also has to be considered.

Table 2. Informational Linkages and Interdependence between Stocks

	Dep. Var.: Monthly Idios. Return		
	(1)	(2)	(3)
	Random Controls	Financial Ratios	Full Controls
$\mathbf{W}^{I}_{t-1}$	0.73	0.72	0.74
	29.32	23.69	30.29
Observations	138,780	138,780	138,780
Controls	Random	Ratios	Full
Stock-Month FE	Yes	Yes	Yes

Notes: The Dep. Var. is the residualized monthly return of stocks after accounting Fama-French 6-Factor model. The first column includes five randomly normally distributed controls. The second includes as controls five financial ratios: Market-Book, Cash-Assets, Ebitda-Assets, Debt-Assets and Log(Sales), all lagged. The third column includes the previously financial ratios controls plus lagged Return and Risk with firm-specific coefficients. T-stats are below each coefficient.

using the lagged matrix in each period to avoid simultaneity.

We use a Spatial Two-Stage Least Squares (S2SLS) instrumental variable procedure, where we instrument the stock returns informationally linked to a stock i with fundamental factors of those linked stocks. To minimize reverse causality concerns, we include various lagged control variables commonly used in finance literature to explain stock returns. Furthermore, we address potential cross-sectional dependence in forecast errors by defactoring analysts' forecast errors, removing factors affecting the entire stock market. This approach also accounts for any unobserved global fundamental factors.

Having proved the relevance of the informational channel, we now move on to discussing different effects that this channel can have in periods with higher uncertainty and considering category learning effects.

5.2. Validity Analysis

We have been trying to show evidence that the proposed informational channel, motivated by learning frictions in agents' learning process that generate correlated beliefs between assets, explains stock market excess comovement. Based on other studies (Barberis and Shleifer, 2003; Barberis et al., 2005; Calvo and Mendoza, 2000; Gorton and Ordoñez, 2014; Israelsen, 2016; Veldkamp, 2006), we perform a validity analysis to understand how our measure of informational interdependence behaves across various scenarios. In this section, we show that our proposed informational channel is still important in explaining stock market excess comovement while considering these different scenarios already studied in the literature.

5.2.1. *Category Learning*

Excess comovement in agents' beliefs (and in stock returns) can also arise due to agents learning in terms of categories. This is studied in the stylized model of Bernales et al. (2024) discussed in our Section 3. The result from this model is based on agents wrongly believing that information signals regarding the common factor are more informative than they are in reality, over-weighting the precision of this signal, due to limited attention capacity or other factors. This assumption is based on previous literature about category learning (Barberis and Shleifer, 2003; Barberis et al., 2005).

The intuition in the literature of category learning is that agents, to simplify their decisions, group assets into categories and allocate funds at the level of these categories. If there are correlated beliefs across investors using these categories, and if their trading affects prices, as funds move from one category to another, there will be a coordinated demand that induces excess comovement. even when these assets' cash flows are uncorrelated (Barberis et al., 2005).

In this section, we analyze if our proposed informational channel is still relevant when considering category-learning effects. To do this, we interact our informational channel matrix, W_t^I with groups that have been proposed in the literature.

One of the most studied groups used in the literature to understand category-learning effects is regarding stocks present or not in the Standard and Poor 500 Index (S&P 500). Barberis et al. (2005) studies changes in market betas of stocks recently added to the S&P 500 and find important effects. Adding a stock to the S&P 500 should not change investors' perceptions of the covariance of the included stock's fundamental value with other stocks' fundamental values, still they find important effects.

We study if firms that have ever been present in the S&P 500 index and are informationally linked show a different intensity of interdependence. Results can be seen in the second column of Table 3, where it can be seen that firms that are informationally linked and have been present in the S&P 500 show a higher intensity of interdependence. The inclusion of this interacting effect does not diminish the positive and significant impact of the informational channel by its own, which decreases in intensity but still shows a large sign of interdependence between stock idiosyncratic returns.

Agents can also naturally group stocks in terms of industries. We test, in Column (4) of Table 3, if firms that belong to the same industry (at the 2-digit NAICS code level) and are informationally linked perceive a different intensity of interdependence. Geographical proximity is also a category in which agents can group stocks. Agents could think that if firms are close to each other, the news will impact them similarly. We consider this group with two measures of geographical proximity, "GeogThres" and "GeogState"[15].

We also consider that stock market excess comovement can arise due to common analyst coverage, as in Israelsen (2016). The intuition is that comovement in the stock market can be driven by a set of analysts following the same stocks. Individual analysts' forecast errors are likely to be correlated across the stocks they cover because they may use the same model or input when making predictions, which will cause any error to propagate through all of those assets.

This idea could raise some concerns about how we estimate informational linkages. We use correlated analyst forecast errors to proxy for shared mistakes in agents' beliefs. A potential concern on our metric is that we could be capturing common analyst coverage.

[15] "GeogThresh" connects firm if their headquarters are at a 100km or less of distance (5% of the firms are at less than 100km of distance). "GeogState" connects firms if they are part of the same state. To give an idea of how securities are located, Figure A.1 shows the location of securities headquarters.

To address this concern, we construct a matrix of common analyst coverage and show that our proposed informational channel is still important even after controlling for common analyst coverage.

We say that a pair of stocks has common analyst coverage if they have had at least one common analyst in the year in question. Since W_t^I uses five years of data to be constructed, we keep this frequency and say that a pair of securities are connected by common analysts if they have had at least one common analyst on two out of the five years in which W_t^I is constructed, we name this matrix $W_t^{ComAnalys}$. To correctly address the critique, instead of just analyzing the interaction between W_t^I and $W_t^{ComAnalys}$, we also include $W_t^{ComAnalys}$ on its own. In the last column of Table 3 we show that the intensity of the relevance of the informational channel does not vary so much when considering $W_t^{ComAnalys}$.

Table 3 shows that the informational channel is still relevant even after considering each of the category-learning effects proposed[16]. It also shows that the intensity of the informational channel increases when considering category-learning effects.

In particular, the second column shows the relevance of the S&P 500 group in amplifying the effects of the informational channel, and the third column of common industry effects. The fourth and fifth show that securities geographically close to each other also amplify the effect.

The last column shows that the informational channel is still relevant even after considering common analyst coverage effects. It also shows that common analyst coverage is relevant in explaining idiosyncratic stock returns, as in Israelsen (2016), but the interaction between

[16]Note that each of the matrices that aim to construct categories are dummy matrices for connections between securities, where an element has a value of one when a pair of securities are connected and zero if not.

Table 3. Informational Linkages and Category Learning

	Dep. Var.: Monthly Idios. Return					
	(1)	(2)	(3)	(4)	(5)	(6)
Groups	Baseline	S&P	Industry	Geog.	Geog.	Common Analy.
W	0.74	0.48	0.35	0.64	0.60	0.63
	30.29	18.27	13.4	25.86	24.54	23.41
W*S&P		0.76				
		13.40				
W*Industry			0.85			
			16.14			
W*GeogThres				9.12		
				3.70		
W*GeogState					12.89	
					4.67	
$W^{ComAnalys}$						0.75
						28.07
$W * W^{ComAnalys}$						6.11
						1.05
Observations	138,780	138,780	138,780	138,780	138,780	138,780
Controls	Yes	Yes	Yes	Yes	Yes	Yes
Stock-Month FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The Dep. Var. is the residualized monthly return of stocks after accounting Fama-French 6-Factor model. Each estimation controls for Market-Book, Cash-Assets, Debt-Assets, Ebitda-Assets, Log(Sales), Return and Risk (all lagged) with firm-specific coefficients, besides stock and monthly FE. Each column includes an interaction for pair of firms that belong to a group. The second column interacts the W matrix keeping only pair of firms that have ever been part of the S&P 500 index. The third column keeps connections between pairs of firms that are in the same industry. The fourth column keeps connections between stoctks hat are at most 100km. distanced (between their headquarters). The fifth column keeps connections that are in the same state. The sixth column builds a matrix of common analyst coverage and then interacts our base W with the common analyst W. T-stats are below each coefficient.

both matrices does not have a significant effect over the dependent variable.

5.2.2. *Uncertainty Periods*

In this section, we analyze how the importance of the informational channel (or learning biases) changes during periods of higher uncertainty. Two strands of literature motivate this analysis: information production and herding behavior.

Gorton and Ordoñez (2014) and Cole et al. (2024) suggest that information acquisition and production change during periods of high uncertainty. In our context, this implies that agents demand and access information differently during such periods, altering their learning biases. These studies suggest that, in our framework, a negative aggregate shock can trigger increased information production about stocks, as agents have more at stake. This could lead to a temporary reduction in learning frictions and, consequently, weaker informational linkages between stocks.

The second strand of the literature focuses on herding and contagion theories (Kaminsky and Schmukler, 1999; Calvo and Mendoza, 2000; Forbes and Rigobon, 2001; Basu, 2002). Calvo and Mendoza (2000) suggest that investors have fewer incentives to gather good news in bullish markets compared to bearish ones. In our context, this implies that the intensity of shock propagation through the informational channel could be lower during periods of higher uncertainty or market stress.

Both strands of literature point towards a potential decrease in the strength of informational linkages during periods of high uncertainty. This leads us to hypothesize that our measure of informational interdependence may exhibit different behavior under such conditions, potentially showing a decrease in intensity.

A particular contribution of our study with respect to most of the herding literature is that this literature tends to study short periods near crises. Instead, we focus on a longer period that includes periods of high and low uncertainty, being able to understand how different is the learning process of individuals during the business cycle.

As in the previous section, we interact our matrix of informational linkages (W^I_{t-1}) with a dummy that is equal to one in periods of higher uncertainty. We use different measures for periods of higher uncertainty. First, we use only the periods related to the Global Financial Crisis (GFC, from July 2007 until March 2009). We also used the CBOE Volatility Index (VIX) and the Economic Policy Uncertainty Index for the United States as variables to measure uncertainty in the market. Based on the distribution of these variables, we selected as periods of high uncertainty those located in the upper 25% and 10% of the distribution (VIXQ75, VIXQ90, Unc. PolQ75 and Unc. PolQ90). Finally, we also considered those periods where the S%P 500 index had the lowest monthly returns, in particular, the periods where the S&P 500 returns were on the lower 25% and 10% of the distribution (S&P Ret.Q25 and S&P Ret.Q10).

Table 4 shows the results of these estimations. As it can be seen, across each specification there is a negative marginal effect over the interaction of W^I_t with periods of higher uncertainty[17]. This means that, just as the theory discussed suggests, there is a decrease in the intensity of the informational channel in periods of higher uncertainty. Intuitively, this can be due to agents knowing that in these periods they have more at stake, and focusing on learning about firms' fundamentals can have a higher reward.

Overall, with each of the applications shown in Tables 2, 3 and 4 it can be seen that the

[17] It is interesting to notice that the amount of connections increases in periods of higher uncertainty, seen as a decrease in the level of the sparseness of the W matrix, but the intensity of these connections decreases in these periods, as it can be seen in the Appendix C, Table A.1.

Table 4. Informational Linkages and Uncertainty Periods

	Dep. Var.: Monthly Idios. Return							
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
	Baseline	GFC	VIXQ75	VIXQ90	Unc. PolQ75	Unc. PolQ90	S&P Ret.Q25	S&P Ret.Q10
W	0.74	0.72	0.71	0.73	0.69	0.72	0.73	0.75
	30.29	27.67	26.43	27.7	25.16	27.00	27.06	28.07
W*Crisis		-0.24						
		-3.08						
W*VIX			-0.15	-0.35				
			-2.9	-4.99				
W*Unc Policy					-0.16	-0.49		
					-3.38	-8.09		
W*S&P Ret							-0.2	-0.56
							-4.31	-9.29
Observations	138,780	138,780	138,780	138,780	138,780	138,780	138,780	138,780
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Stock-Month FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Notes: The Dep. Var. is the residualized monthly return of stocks after accounting Fama-French 6-Factor model. Each estimation controls for Market-Book, Cash-Assets, Debt-Assets, Ebitda-Assets, Log(Sales), Return and Risk (all lagged) with firm-specific coefficients, besides stock and monthly FE. Each column includes an interaction for crises periods. The second column, Global Financial Crisis (GFC) considers 2007m7-2009m3. VIXQ75 considers periods where the VIX index distribution was in its upper 25%, same for VIXQ90, but with the upper 10%. Unc. Pol. uses the Uncertainty Policy Index. S&P Ret uses periods where the monthly return of the S&P index was in its lower 25% and 10% values. T-stats are below each coefficient.

informational channel can explain idiosyncratic stock returns. Thus, showing interdependence between securities, which in the end generates excess comovement between stock returns.

5.3. Robustness

In this section, we focus on showing how robust are our results to different choices that we had to make regarding the methodology used. We focus on the results provided in the first subsection of the Results section, specifically, results shown in Table 2 in Section 5.1.

First, we had to choose an error measure for analysts' forecast of Earnings Per Share (EPS). We chose the Normalized Mean Root Squared Error (NMRSE). Normalization can be of particular use when considering that different stocks can have different levels of EPS, normalization makes this number comparable across stocks.

We use different measures for forecast errors in columns (1) and (2) of Table 5. In particular, the Normalized Simple Difference (NSD) uses the Simple Difference between forecasted and observed EPS instead of the Root Mean Squared Error. This measure is also normalized by the level of the variable observed. The second measure of forecast error we use is the Root Mean Square Error (RMSE), without normalizing. We find that the error measure that we use is not qualitatively relevant for our baseline result regarding the importance of informational linkages in explaining interdependence between stocks.

We also check if using a dynamic matrix of informational linkages plays a role in explaining our results. We argue that these connections between stocks can change over time, but it also could be argued that we want to capture persistent connections. We use a fixed W matrix to capture informational linkages constructed between 2000-2004, before our regression sample starts. We find that the intensity of interdependence under a fixed W matrix is still important and even higher than what we found in Table 2. Finally, Column (4) changes the threshold with which we define a connection between a pair of stocks as relevant to $p = 0.05$. This means that we allow for a higher amount of connections between pairs of stocks, since our baseline result sets that level in $p = 0.01$. Results show that the result is even higher and still strong.

Table 5. Robustness Checks

	Dep. Var.: Monthly Idios. Return			
	(1)	(2)	(3)	(4)
	RECM	NSD	Fix W	$p = 0.05$
W	0.65	0.86	0.96	0.86
	27.11	34.26	32.49	32.00
Observations	138,780	138,780	138,780	138,780
Controls	Yes	Yes	Yes	Yes
Stock-Month FE	Yes	Yes	Yes	Yes

Notes: The Dep. Var. is the residualized monthly return of stocks after accounting Fama-French 6-Factor model. Each estimation controls for Market-Book, Cash-Assets, Debt-Assets, Ebitda-Assets, Log(Sales), Return and Risk (all lagged) with firm-specific coefficients, besides stock and monthly FE. (1) uses the Root Mean Squared Error as a forecast error measure, (2) uses the Normalized Simple Difference as a forecast error measure. (3) uses a fix instead of dynamic W matrix to account for informational linkages. (4) sets the threshold to define informational linkages strong enough to be 0.05 instead of 0.01 T-stats are below each coefficient.

6. Applications

Having shown the relevance of the informational channel, we now move to applications that we can do with our baseline model. Since we estimated a network of linkages at a stock-to-stock level, we can easily move to analyzing how idiosyncratic shocks to a firm, industry, or state propagate through the stock market. As it has been deeply studied, for instance in Acemoglu et al. (2012), idiosyncratic shocks to individual units (e.g., stocks) can generate aggregate fluctuations when network effects are relevant.

We analyze how simulated and real-world idiosyncratic shocks propagate through the stock market. We start with a simulated shock to each stock (one by one, separately) and analyze the average indirect impact over the stock market. Then, based on a climate event (Hurricane Florence, which is a real-world shock) we analyze if our estimated network captures indirect effects over the stock market.

Our baseline model allows us to identify the direct and indirect impacts of a shock over one stock. The model can be written as[18]:

$$\mathbf{r}_t = (I_N - \rho \mathbf{W}^I_{t-1})^{-1}(\mathbf{X}_{t-1}\beta_i + \boldsymbol{\mu} + \gamma_t + \mathbf{v}_t)$$

The impact of a shock over the error term can be seen as:

$$\frac{\partial \mathbf{r}_t}{\partial \mathbf{v}_t} = (I_N - \rho \mathbf{W}^I_{t-1})^{-1} \tag{6}$$

Note that if we apply a shock in one period ($\mathbf{v}_t$), the resulting derivative will give us an $N \times N$ matrix. The diagonal of the matrix has the direct effects of a shock over each firm. The non-diagonal elements will include the indirect effects of the shock, thus, these elements show the network propagation of the shock.

Direct effects are those that impact directly the firm. For example, if we set a one standard deviation of their idiosyncratic returns shock to each stock, the direct effect of the shock to the same firm is going to be one standard deviation of their idiosyncratic return. Indirect effects are the effects that come from the propagation of the shock through the network. A shock over security "A" will also affect those securities that are connected to security "A". But shocks are not propagated only through first-order connections in the network, if

[18]If $\rho < 1$, $(I_N - \rho \mathbf{W}^I_{t-1}\mathbf{r}_t)^{-1}$ can be written as: $(I_N + \rho W^I_{t-1} + \rho^2 {W^I_{t-1}}^2 + ...)$. The first element includes the direct effect, the second one the impact over direct "neighbors", the third one the impact over second-order "neighbors", and so on.

stock “A” is connected to stock “B” and stock “B” is connected to stock “C”, then a shock over stock “A” can also affect stock “C”. This is a second order connection.

From the previous paragraph, it can directly be inferred that the propagation of a shock is highly dependent on how connected a stock is. More connected stocks are those that are more central. We analyze how heterogeneous are the effects of shocks depending on how central a security is. This gives insights into systemic risk analysis, negative shocks over some securities (more central ones) can be a higher risk for the stock market's performance.

In a financial network, a stock is considered central if it has significant influence over other stocks due to its connections. PageRank centrality is a useful measure for identifying the importance of nodes within a network. It is similar to a highly used measure in the network literature, EigenCentrality, the main difference is that PageRank accounts for link direction. Each node in a network is assigned a score based on its number of incoming links (its 'indegree'). These links are also weighted depending on the relative score of its originating node. We measure the centrality of each stock based on its PageRank centrality in the informational linkages matrix.

6.1. Simulated Shocks

We set a shock to each stock i equal to one-standard deviation of its idiosyncratic returns (the avg. across firms is 8.86%), and estimate its indirect impacts over $j \neq i$, as in Equation (6)[19]. This process is repeated for every stock in our sample, creating a series of shock scenarios. For each scenario, we calculate the average percentage change in idiosyncratic returns across all non-shocked stocks, relative to the mean idiosyncratic return of the entire

[19] Since W^I_{t-1} changes each year, we perform this exercise for every year and report the average.

cross-section. This relative indirect effect is performed using the following equation:

$$\text{Relative Indirect Effect} = \frac{1}{N-1}\sum_{j\neq i}\frac{\Delta r_{j,t}}{\bar{r}}, \tag{7}$$

where N is the total number of stocks, i is the shocked stock, j represents all other stocks, $\Delta r_{j,t}$ is the change in idiosyncratic return of stock j due to the shock on stock i, $\bar{r}$ is the mean idiosyncratic return of the cross-section of stocks.

Table 6 shows the results of this exercise. The first column shows that on average, a one standard deviation shock over i has a relative indirect effect of 14.40% over $j \neq i$. Note that this does not mean that the shock implies a 14.40% increase in the idiosyncratic return of stocks $j \neq i$, but it implies that the shock generates an important increase in the idiosyncratic returns of firms $j \neq i$, relative to the mean idiosyncratic return.

Table 6. Indirect Impact of a 1 sd shock

	Avg. Indirect Impact of a Shock to:			
	(1)	(2)	(3)	(4)
	All Firms	More Connected	Least Connected	First Order
Relative Indirect Effect	14.40%	32.65%	0.73%	28.54%

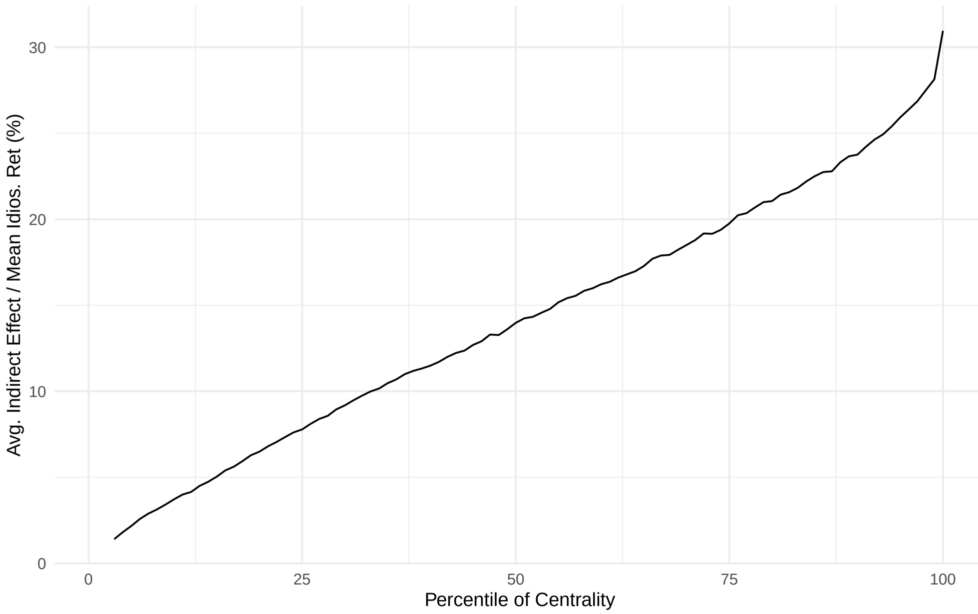
Notes: Estimates correspond to the relative indirect effects of a one standard-deviation exogenous shock to a given security i's idiosyncratic return (on avg. 8.86%) on other securities $j \neq i$. The first column applies the shock to all firms (separately) and reports the average across shocked firms (as in all columns). The second column applies the shock to the 5 most connected stocks according to Pagerank centrality. The third column applies the shock to the 5 least connected stocks according to Pagerank centrality, conditional on having at least one connection. The last column applies the shock to all firms but estimates the indirect effect only over 1st-order connections, divided by the number of 1st-order connections. Results are obtained from 1000 simulated coefficients drawn from the multivariate normal distribution implied by the estimated variance-covariance matrix obtained from the estimates of the baseline estimation (last column of Table 2).

Columns (2) and (3) separate the effects according to the centrality of firms. Column (2) applies the shock only to the five most connected stocks. A shock to these highly connected firms results in a relative indirect effect of 32.65% indicating that shocks to central firms have more than twice the effects of an average firm. Column (3) focuses on the five least connected firms, conditional to them having at least one connection. The relative indirect effect here is only 0.73%, showing that shocks to peripheral firms have minimal effects on the network.

Column (4) is similar to (1), but it analyzes only the first-order impact of the shock. It applies the shock to one firm i, sums the indirect impact over firms $j \neq i$ that are directly (first order) connected to i and *divides only across connected firms*. The difference with column (1) is that (1) sums over all $j \neq i$, even if they are not directly connected, divides over all $N - 1$ firms, and takes the average across all stocks. Since second and higher-order connections propagate the shock with less intensity, it is expected for Column (4) to have a higher value than Column (1). This also shows that first-order connections propagate most of the shock.

Figure 1 extends the analysis from Columns (2) and (3) and plots the average firm-level indirect effects by percentile of centrality. As we have already mentioned, more central firms propagate more idiosyncratic shocks than less central firms. How relevant is this difference? In the figure it can be observed that firms with higher centrality (closer to the end of the X-axis) propagate shocks disproportionately more than firms with lower centrality, there is a clear change in slope on the right side of the plot. This means that a subset of firms propagates shocks with a much higher intensity than the rest.

Figure 1. Relative Indirect Effect of a 1 sd shock (8.86%) by Quantiles of Centrality



Notes: Quantiles of centrality are calculated following PageRank centrality and the baseline matrix of connections, W^I. Aggregate indirect effects are simulated based on a 1 standard deviation shock to a firm i. Results are obtained from 1000 simulated coefficients drawn from the multivariate normal distribution implied by the estimated variance-covariance matrix obtained from the estimates of the baseline estimation (last column of Table 2).

6.2. Real-World Shocks

In this section, we take a climate event, Hurricane Florence, and analyze up to what point our estimated network can capture the network indirect effects of a real-world shock. Hurricane Florence was a powerful and long-lived hurricane that caused widespread damage in North Carolina (NC), South Carolina (SC), and Virginia (VA) in September 2018. It affected over 1.5M people, it had uninsured damages for $9M USD and 53 people died. It was particularly notable for its slow movement and intense rainfall, leading to severe flooding and significant economic disruption. We examine the stock market's response to this event.

Our approach is simple; we analyze the stock market's response to firms affected by the shock: firms with their headquarters (HQ) in NC, SC, and VA. Then, we analyze the stock market response to firms not shocked (with headquarters that are not in NC, SC, and VA), which would be the indirect effects. Finally, we analyze the stock market response of firms that are connected and not in an affected state, according to *our* matrix of informational linkages, to firms in the affected states.

Table 7 shows that the average return on September 2018 of stocks with headquarters in affected states was -1.55%. The mean return of our subset of stocks (771) in that month was -0.34%, the mean return of stocks that were not affected (HQ not in NC, SC, or VA) was (-0.26%/-1.55%) 17.13% of the mean return of affected stocks. But the mean return of stocks that are informationally linked to stocks in NC, SC, or VA[20], was (-0.59%/-1.55%) 37.97% of the affected stocks.

Table 7. Hurricane Florence Indirect Effects

	(1) Affected	(2) 1st Order	(3) Not Affected	(4) All
Avg. Return	-1.55%	-0.59%	-0.26%	-0.34%
Avg. Return/Affected	100 %	37.97%	17.13%	22.00%

Notes: The first row calculates the average monthly return in September 2018 for different sets of stocks. (1) does it for stocks with HQ in SC, NC, and VA. (2) does it for stocks informationally linked to stocks in (1), but not present in SC, NC, and VA. (3) does it for all stocks with HQ not in SC, NC, or VA. (4) does it for all stocks. The second row divides the first row results over the average monthly return in Sept. 2018 for stocks affected (-1.5%).

One would expect the shock to propagate with more intensity over stocks directly linked

[20] This was done using W^I_{2018}, which is constructed using data between 2013 and 2017.

to those affected stocks. Capturing those links is not obvious, in this exercise, we show that our estimated matrix of informational connections can capture the propagation of real-world shocks. It is possible that firms that have their headquarters in a state different from the affected stocks still have a presence in those states, and thus, are exposed to the shock. We cannot rule out this effect, and thus we show a lower bound of the indirect impacts of the shock.

7. Conclusion

This paper provides empirical evidence that excess comovement in stock returns can be explained by stock-to-stock interdependence through informational links. We contribute to the literature on stock market excess comovement by proposing that analysts' learning biases can lead to correlated beliefs between agents, which we term informational linkages. We provide evidence that these linkages generate interdependence between stock returns at a micro-level and contribute to stock market comovement at an aggregate level.

Our novel measure for informational linkages, based on analysts' forecast errors, captures the learning process of agents, which is biased towards common information due to learning frictions. We show that these informational linkages explain stock returns even after controlling for fundamental connections and considering various explanations already studied in the literature.

We provide evidence that our informational channel is still relevant after taking into account different channels already studied in the literature, such as agents learning in terms of categories and grouping stocks based on categories. We also study the dynamics of these potential learning biases. During periods of higher uncertainty, we observe a decrease in the intensity of informational linkages.

To illustrate the practical implications of our findings, we apply our estimated matrix of informational linkages to both simulated and real-world shocks. Our simulations reveal that idiosyncratic shocks to individual firms can have significant propagation effects throughout the stock market, with more central firms disproportionately amplifying these effects. Furthermore, we use a real-world climate event, demonstrating that our estimated informational linkage matrix can effectively capture the actual propagation of shocks in the market.

Our channel is based on biases in the learning process of agents. If agents understand their biases, they can try to improve their decisions and create a more efficient environment in financial markets. Understanding the role of informational links in shaping the dynamics of stock returns is also of huge relevance for investors in making portfolio decisions and risk management. This analysis allows us to understand in a better sense how shocks are propagated through the stock market. Knowing how central is a firm or industry in the matrix of connections is important to understand the effects that a shock will have.

References

Acemoglu, Daron, Vasco M. Carvalho, Asuman Ozdaglar, and Alireza Tahbaz-Salehi. 2012. "The Network Origins of Aggregate Fluctuations." *Econometrica* 80 (5): 1977–2016. https://doi.org/10.3982/ECTA9623.

Antón, Miguel, and Christopher Polk. 2014. "Connected Stocks." *The Journal of Finance* 69 (3): 1099–1127. https://doi.org/10.1111/jofi.12149.

Bailey, Natalia, Sean Holly, and M. Hashem Pesaran. 2016. "A Two-Stage Approach to Spatio-Temporal Analysis with Strong and Weak Cross-Sectional Dependence." *Journal of Applied Econometrics* 31 (1): 249–280. https://doi.org/10.1002/jae.2468.

Barberis, Nicholas, and Andrei Shleifer. 2003. "Style investing." *Journal of Financial Economics* 68 (2): 161–199, https://EconPapers.repec.org/RePEc:eee:jfinec:v:68:y:2003:i:2:p:161-199.

Barberis, Nicholas, Andrei Shleifer, and Jeffrey Wurgler. 2005. "Comovement." *Journal of Financial Economics* 75 (2): 283–317. https://doi.org/10.1016/j.jfineco.2004.04.003.

Basu, Ritu. 2002. "Financial Contagion and Investor "Learning"': An Empirical Investigation." *IMF Working Paper* (No. 2002/218): .

Bekaert, Geert, Michael Erhmann, Marcel Fratzscher, and Arnaud Mehl. 2014. "The Global Crisis and Equity Market Contagion." *The Journal of Finance* 69 (6): 2597–2649, http://www.jstor.org/stable/43611077.

Bernales, Alejandro, Hriday Karnani, and Paula Margaretic. 2024. "Informational

Economic Transmission between Countries." https://dx.doi.org/10.2139/ssrn.4801634.

Calvo, Guillermo A., and Enrique G. Mendoza. 2000. "Rational contagion and the globalization of securities markets." *Journal of International Economics* 51 (1): 79–113. https://doi.org/10.1016/S0022-1996(99)00038-0.

Chousakos, Kyriakos, Gary Gorton, and Guillermo Ordoñez. 2023. "Information Dynamics and Macro Fluctuations." *American Economic Journal: Macroeconomics* 15 (4): 372–400. 10.1257/mac.20210101.

Coibion, Olivier, and Yuriy Gorodnichenko. 2015. "Information Rigidity and the Expectations Formation Process: A Simple Framework and New Facts." *The American Economic Review* 105 (8): 2644–2678, http://www.jstor.org/stable/43821351.

Coibion, Olivier, Yuriy Gorodnichenko, and Saten Kumar. 2018. "How Do Firms Form Their Expectations? New Survey Evidence." *The American Economic Review* 108 (9): 2671–2713, https://www.jstor.org/stable/26528539.

Coibion, Olivier, Yuriy Gorodnichenko, and Tiziano Ropele. 2019. "Inflation Expectations and Firm Decisions: New Causal Evidence*." *The Quarterly Journal of Economics* 135 (1): 165–219. 10.1093/qje/qjz029.

Cole, Harold, Daniel Neuhann, and Guillermo Ordoñez. 2024. "Information Spillovers and Sovereign Debt: Theory Meets the Eurozone Crisis." *Review of Economic Studies*.

Colla, Paolo, and Antonio Mele. 2010. "Information Linkages and Correlated Trading." *The Review of Financial Studies* 23 (1): 203–246, http://www.jstor.org/stable/40468307.

David, Joel M., and Ina Simonovska. 2016. “Correlated beliefs, returns, and stock market volatility.” *Journal of International Economics* 99 S58–S77. https://doi.org/10.1016/j.jinteco.2015.11.006, 38th Annual NBER International Seminar on Macroeconomics.

Debarsy, Nicolas, Cyrille Dossougoin, Cem Ertur, and Jean-Yves Gnabo. 2018. “Measuring sovereign risk spillovers and assessing the role of transmission channels: A spatial econometrics approach.” *Journal of Economic Dynamics and Control* 87 21–45. https://doi.org/10.1016/j.jedc.2017.11.005.

Elhorst, J. Paul, Marco Gross, and Eugen Tereanu. 2021. “Cross-Sectional Dependence and Spillovers in Space and Time: Where Spatial Econometrics and Global VAR Models meet.” *Journal of Economic Surveys* 35 (1): 192–226. https://doi.org/10.1111/joes.12391.

Fama, Eugene F, and Kenneth R French. 1993. “Common risk factors in the returns on stocks and bonds.” *Journal of Financial Economics* 33 (1): 3–56.

Fama, Eugene F, and Kenneth R French. 2018. “Choosing factors.” *Journal of Financial Economics* 128 (2): 234–252.

Forbes, Kristin, and Roberto Rigobon. 2001. *Measuring Contagion: Conceptual and Empirical Issues*. 43–66, Boston, MA: Springer US, . 10.1007/978-1-4757-3314-3_3.

Froot, Kenneth A., David S. Scharfstein, and Jeremy C. Stein. 1992. “Herd on the Street: Informational Inefficiencies in a Market with Short-Term Speculation.” *Journal of Finance* XLVII (4): 1461–1484.

Giglio, Stefano, Matteo Maggiori, Johannes Stroebel, and Stephen Utkus. 2021. “Five Facts about Beliefs and Portfolios.” *American Economic Review* 111 (5): 1481–1522. 10.1257/aer.20200243.

Gorton, Gary, and Guillermo Ordoñez. 2014. "Collateral Crises." *American Economic Review* 104 (2): 343–78. 10.1257/aer.104.2.343.

Greenwood, Robin. 2007. "Excess Comovement of Stock Returns: Evidence from Cross-Sectional Variation in Nikkei 225 Weights." *The Review of Financial Studies* 21 (3): 1153–1186. 10.1093/rfs/hhm052.

Hameed, Allaudeen, Randall Morck, Jianfeng Shen, and Bernard Yeung. 2015. "Information, Analysts, and Stock Return Comovement." *The Review of Financial Studies* 28 (11): 3153–3187. 10.1093/rfs/hhv042.

Holm, Sture. 1979. "A Simple Sequentially Rejective Multiple Test Procedure." *Scandinavian Journal of Statistics* 6 (2): 65–70, http://www.jstor.org/stable/4615733.

Israelsen, Ryan D. 2016. "Does Common Analyst Coverage Explain Excess Comovement?" *The Journal of Financial and Quantitative Analysis* 51 (4): 1193–1229, http://www.jstor.org/stable/44157611.

Kaminsky, Graciela, and Sergio Schmukler. 1999. "What Triggers Market Jitters? A Chronicle of the Asian Crisis." *World Bank Working Paper* (No. 2094): .

Kelejian, Harry H, and Ingmar Prucha. 1998. "A Generalized Spatial Two-Stage Least Squares Procedure for Estimating a Spatial Autoregressive Model with Autoregressive Disturbances." *The Journal of Real Estate Finance and Economics* 17 (1): 99–121, https://EconPapers.repec.org/RePEc:kap:jrefec:v:17:y:1998:i:1:p:99-121.

Kelejian, Harry H., and Ingmar Prucha. 2004. "Estimation of simultaneous systems of spatially interrelated cross sectional equations." *Journal of Econometrics* 118 (1-2): 27–50, https://EconPapers.repec.org/RePEc:eee:econom:v:118:y:2004:i:1-2:p:27-50.

Kelejian, Harry H., and Ingmar R. Prucha. 1999. "A Generalized Moments Estimator for the Autoregressive Parameter in a Spatial Model." *International Economic Review* 40 (2): 509–533. https://doi.org/10.1111/1468-2354.00027.

Kumar, Alok, Jeremy K. Page, and Oliver G. Spalt. 2016. "Gambling and Comovement." *Journal of Financial and Quantitative Analysis* 51 (1): 85–111. 10.1017/S0022109016000089.

Leary, Mark T., and Michael Roberts. 2014. "Do Peer Firms Affect Corporate Financial Policy?" *Journal of Finance* 69 (1): 139–178, https://EconPapers.repec.org/RePEc:bla:jfinan:v:69:y:2014:i:1:p:139-178.

Lee, Charles M. C., Andrei Shleifer, and Richard H. Thaler. 1991. "Investor Sentiment and the Closed-End Fund Puzzle." *The Journal of Finance* 46 (1): 75–109, http://www.jstor.org/stable/2328690.

LeSage, James P., and Christine Thomas-Agnan. 2015. "Interpreting Spatial Econometric Origin-Destination Flow Models." *Journal of Regional Science* 55 (2): 188–208, https://ideas.repec.org/a/bla/jregsc/v55y2015i2p188-208.html.

LeSage, James, and Robert K.. Pace. 2009. *Introduction to Spatial Econometrics*. London New York: CRC Press.

Muslu, Volkan, Michael Rebello, and Yexiao Xu. 2014. "Sell-Side Analyst Research and Stock Comovement." *Journal of Accounting Research* 52 (4): 911–954. https://doi.org/10.1111/1475-679X.12057.

Peng, L. 2005. "Learning with information capacity constraints." *Journal of Financial and Quantitative Analysis* 40 307–330.

Peng, Lin, and Wei Xiong. 2006. "Investor attention, overconfidence and category learning." *Journal of Financial Economics* 80 (3): 563–602. https://doi.org/10.1016/j.jfineco.2005.05.003.

Pindyck, Robert S., and Julio J. Rotemberg. 1993. "The Comovement of Stock Prices*." *The Quarterly Journal of Economics* 108 (4): 1073–1104. 10.2307/2118460.

Scharfstein, David S., and Jeremy C. Stein. 1990. "Herd Behavior and Investment." *Amercian Economic Review* 80 (Jun.): 465–479.

Veldkamp, Laura L. 2006. "Information Markets and the Comovement of Asset Prices." *The Review of Economic Studies* 73 (3): 823–845, http://www.jstor.org/stable/20185030.

Appendix

A. Extracting the Idiosyncratic Component of Stock Returns

We extract the idiosyncratic component of stock returns by following the steps detailed below:

1. Compute the excess return (**er**) of each asset by subtracting the effective daily return of the asset by a risk-free instrument.

2. At the asset level, keep idiosyncratic returns by de-factoring the asset excess returns with Fama-French's factors.

$$\mathbf{er}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{F}_t + \mathbf{u}_t \tag{8}$$

Excess returns are denoted by $\mathbf{er}_t$, $\mathbf{F}_t$ are Fama-French's 6 factors: (i) Excess Return on the Market (Mkt-Rf) ; (ii) Small Minus Big (SMB); (iii) High Minus Low (HML); (iv) Robust Minus Weak (RMK); (v) Conservtive Minus Aggresive (CMA); (vi) Momentum (Mom) [21]. Note that **ff** is a vector of $N \times 1$ stock-specific slopes and **fi** is a matrix of $N \times 6$ coefficients for each firm-factor.

We keep the residuals, $\hat{\mathbf{u}}_t$, from the OLS estimation of Equation (8), which we call the

[21](i) Mkt-Rf comes from $(r_m - r_f)$, where r_m is the return of the market. (ii) SMB is the average return on the nine small stock portfolios minus the average return on the nine big stock portfolios. (iii) HML is the average return on the two value portfolios minus the average return on the two growth portfolios. (iv) RMW is the average return on the two robust operating profitability portfolios minus the average return on the two weak operating profitability portfolios. (v) is the average return on the two conservative investment portfolios minus the average return on the two aggressive investment portfolios. (vi) Mom is the average return on the two high prior return portfolios minus the average return on the two low prior return portfolios. More information on this link.

idiosyncratic returns of a security. In the following equations, $\hat{\mathbf{u}}_t = \mathbf{r}_t$.

3. Finally, we aggregate daily observations at a monthly frequency: $\mathbf{r}_t^{Monthly} = \prod_{t \in T^{DiM}}(1 + \mathbf{r}_t^{Daily}) - 1$. Where T^{DiM} is the number of days in each month.

B. Spatial Two-Stage Least Squares (S2SLS) Details

We follow and adapt Kelejian and Prucha (1998) S2SLS procedure. We consider the following instruments:

$$\mathbf{H_n} = \left[\left\{ \left(\mathbf{W}_{t-1}^{I} \otimes \mathbf{I_T} \right) \mathbf{X}_{t-1} \right\}_{t \in \mathbf{T}}, \mathbf{X}_{t-1} \right]$$

which allows the estimation of the following parameters

$$\widehat{\theta} = \begin{bmatrix} \widehat{\rho} \\ \widehat{\beta}_i \end{bmatrix} = \left[\mathbf{Z}'\mathbf{Q}\mathbf{H_n} \left(\mathbf{H}_\mathbf{n}'\mathbf{Q}\mathbf{H_n} \right)^{-1} \mathbf{H}_\mathbf{n}'\mathbf{Q}\mathbf{Z} \right]^{-1} \mathbf{Z}'\mathbf{Q}\mathbf{H_n} \left(\mathbf{H}_\mathbf{n}'\mathbf{Q}\mathbf{H_n} \right)^{-1} \mathbf{H}_\mathbf{n}'\mathbf{Q}\mathbf{y},$$

where $\mathbf{Z}$ denotes the matrix of all LHS variables (including the spatial lags of the dependent variables) and $\mathbf{Q}$ is a matrix that sweeps all fixed effects and the intercept, that is,

$$\mathbf{Q} = \left(\mathbf{I_N} - \mathbf{N}^{-1}\mathbf{J_N} \right) \otimes \left(\mathbf{I_T} - \mathbf{T}^{-1}\mathbf{J_T} \right),$$

with $\mathbf{J}$ a square matrix of ones. The standard errors can be obtained as

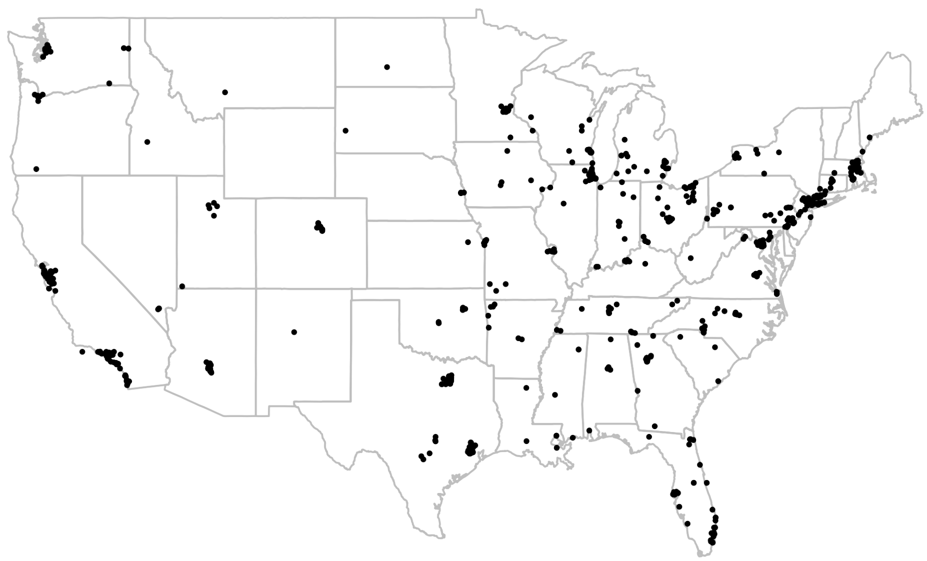
$$\mathrm{se}(\widehat{\theta}) = \widehat{\sigma}^2 \left[\mathbf{Z}'\mathbf{Q}\mathbf{H_n} \left(\mathbf{H}_\mathbf{n}'\mathbf{Q}\mathbf{H_n} \right)^{-1} \mathbf{H}_\mathbf{n}'\mathbf{Q}\mathbf{Z} \right]^{-1}$$

with

$$\hat{\sigma}^2 = \frac{\hat{\mathbf{v}}'\mathbf{Q}\hat{\mathbf{v}}}{(N-1)(T-1)-(k+T)}$$

C. Additional Figures and Tables

Figure A.1. Securities Headquarters



Notes: Each dot represents the location of the Headquarters (based on the Zipcode) of each stock used in our analysis. Own elaboration based on data from Compustat.

Table A.1. Level of Sparsness and Intensity of W by year

Year	Sparsness	Intensity
2005	80.19%	0.67%
2006	80.64%	0.68%
2007	84.76%	0.87%
2008	84.08%	0.82%
2009	80.17%	0.66%
2010	81.63%	0.71%
2011	82.45%	0.74%
2012	82.54%	0.75%
2013	78.16%	0.59%
2014	89.15%	1.19%
2015	90.74%	1.41%
2016	90.91%	1.43%
2017	91.22%	1.48%
2018	91.52%	1.53%
2019	91.58%	1.52%

Note: Sparsness is the percentage of not-connected stocks (0s) in our informational interdependence matrix for each year, as described in Section 3. Intensity is the degree average degree of intensity of the links in our informational interdependence matrix each year.