Information Diffusion, Interdependence and Stock Market Comovement*

Hriday Karnani†

Master in Economics Thesis

Advisors: Alejandro Bernales, Paula Margaretic and Eugenia Andreasen

This draft: July 21, 2024

Abstract

We explore how comovement in stock returns can be explained by firm-to-firm interdependence through informational links. Due to limited attention capacity, agents' learning process is biased towards common information, which generates correlated beliefs about stock returns. This will generate interdependence across stock returns, and stock market comovement at an aggregate level. We use a novel measure for informational linkages between firms based on analysts' forecast errors, estimate the relevance of this channel with a Spatial Two-Stage Least Square (S2SLS) estimator, and find that the informational channel explains "idiosyncratic" returns, has a lower intensity in periods of higher uncertainty, and is amplified when taking into account category-learning effects. We also study the propagation of climate events and simulated shocks in the stock market based on the estimated informational linkages and find quantitatively important indirect effects.

JEL Codes: D80, G12, G14.

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

---

*I am extremely grateful to Alejandro Bernales, Paula Margaretic and Olga Kolokova for their invaluable guide, time, support and patience in this process. I also thank Eugenia Andreasen and Jorge Miranda-Pinto 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

The asset pricing literature has extensively studied what determines asset prices. It has also been widely documented that some assets, such as stock, perceive a high comovement between their returns (Froot and Dabora, 1999; Barberis et al., 2005; Veldkamp, 2006) even after controlling for common factors. This phenomenon has been attributed to different factors in the literature, such as contagion, herding, interdependence, among others.

The topic can be seen as increasingly important due to the recent crises we have suffered. Understanding how the effects of a shock on an asset could affect other stocks and the market as an aggregate is crucial for policymakers and practitioners. Stock return comovement is also vital for assessing risk, portfolio management, testing the efficiency of markets, among many other topics.

Stock prices are determined by many factors; between them, fundamentals of stocks and factors related to the state of the economy are key. Stock returns can be divided into these two components: an idiosyncratic -specific to each firm-, and an aggregate or common factor -regarding a group of the economy. If we separate these two components, by definition the idiosyncratic component should not be highly correlated with other stocks[1]. What we observe in the data is that there is a comovement between the idiosyncratic component of stock returns, as can be seen in Figure 1.

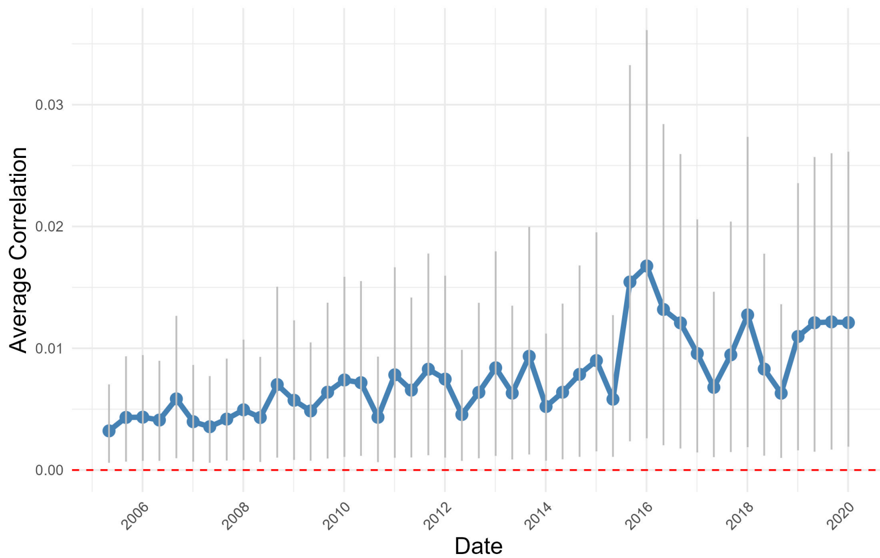
Other factors, aside of fundamentals and factors related to the state of the economy, are also relevant, such as agents' learning process about information related to the assets, the way they form expectations (beliefs) and the diffusion of information available in the market.

In this paper, we explore how learning biases in agents' learning process can induce correlated beliefs across stocks, which in the end could generate comovement in the stock market as an aggregate. We do this by constructing stock-to-stock *informational* linkages and check if these linkages are able to show signs of interdependence between stocks, and comovement at an aggregate level.

Investors trade stocks based on their beliefs, which depend on the information they consume (Giglio et al.,

[1]Throughout this paper we will focus on the "idiosyncratic" component of stock returns. I will use the term "idiosyncratic". People might asume, correctly, that when someone uses this term, there does not exist any comovement between different units. In this paper we will use the term idiosyncratic and still allow for comovement between different units. This is because when we refer to idiosyncratic, we refer to the separation between this component and the aggregate one in stock returns, as suggested by basic financial theory.

Figure 1: Avg. Corr. of Idios. Daily Returns



Notes: This figure plots the average correlation between monthly stock returns of pairs of stocks every four months. Gray bars represent confidence intervals at the 95% level. Source: Own elaboration based on data from CRSP.

2021). We have seen an increase in the amount of information available, which also requires a higher capacity to consume more information. It has been widely studied that investors have limited attention capacity and it is costly for them to learn about new information -in terms of time and money- (Veldkamp, 2006), this can induce the learning process of agents' to have frictions, specifically, to be biased towards consuming or emphasizing more certain type of information. These learning biases can generate correlated errors in the beliefs of agents.

A simple example can illustrate the idea of the last paragraphs. Think of an agent that trades in the stock market. This agent wants to learn about each potential stock in which they will invest. They search for information about stocks and consume two types of information: (i) idiosyncratic information about one stock and (ii) common information about a group of stocks. For simplicity, let us take the common information about a group of stocks to be about pair of stocks[2].

Agents have to interpret the information they consume, which they do imperfectly. The degree of precision with which they interpret the information will depend on the effort they put into interpreting each type of

[2]To be more complete, we could add a (iii) component to be information about the whole economy. We ignore this component for simplicity.

information [(i) or (ii)]. Since agents have a limited attention capacity and they face learning costs (frictions in their learning process), they will focus more on information about a pair of stocks, since it gives them information about more stocks in less time and money. But by doing this, there beliefs about these pair of stocks are going to be correlated, that will induce them to invest in a similar way between those two stocks, even when those two stocks might have completely different fundamentals. This can generate interdependence between stocks, and comovement in the stock market as an aggregate. In Section 3.1 we explain with more depth and formality the intuition given here.

We construct a matrix of informational linkages between pairs of stocks based on analyst forecast errors, which we assume to be a proxy for beliefs. Then, we test the relevance of these informational links in explaining interdependence between stocks, following a Spatial Two-Stages Least Squares (S2SLS) methodology. We show that the informational channel proposed is able to explain "idiosyncratic" returns of stocks, which means that there is evidence of interdependence between the idiosyncratic component of stock returns. This interdependence might explain stock market comovement at an aggregate level if the direction of this interdependence tends to be the same across stocks, which we observe.

We find that informational linkages behave differently on periods with higher uncertainty. In particular, the intensity of these channels decreases. This is consistent with some theories (Calvo and Mendoza, 2000; Gorton and Ordoñez, 2014) that focus on herding in bullish and bearish markets, and the optimal amount of information production in periods with greater uncertainty.

We also consider the strand of literature that tries to explain comovement based on category-learning theories (Barberis and Shleifer, 2003; Barberis et al., 2005). It could be argued that the linkages we find are just groups in which people categorize stocks. We find that our result stands even after considering these different categories, such as stocks that are part of the S&P 500, that are in the same industry, that are geographically close and that share common analysts coverage.

We then move into applications where we use the linkages estimated. We use our informational linkages matrix between stocks and estimate the propagation (indirect effects) through the stock market of simulated and real world shocks to one (or a small amount) of stocks. We find evidence that indirect effects of these shocks -the effect of a shock over stocks not directly affected by it-, are quantitatively important.

Keeping on with the analysis of shock propagation, we also study how heterogeneous is the previous result depending on how central each stock is. More central firms are those that are more connected[3]. Those stocks will propagate shocks more than others, and we quantify this effect. We also take climate events that affected one or a few states, and show how well our informational linkages matrix captures the real indirect effects of the shock. We find that the indirect effects of these climate events are quantitatively important and can be partially captured by our estimated matrix of connections.

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 in more detail the informational channel, our empirical strategy, and explores potential drivers of heterogeneity in our results. Section 4 includes the results and the implications that come out of them. Section 5 studies the simulation and real-world propagation of shocks through our network, and Section 6 concludes.

2 Literature Review

We contribute to the literature studying agents' learning process, their beliefs, and how both of them can generate certain types of trading behavior. Previous literature, such as Coibion and Gorodnichenko (2015) provides empirical evidence suggesting that professional forecasters do not fully adhere to the traditional full-information hypothesis of rational expectations. Instead, they form expectations while constrained by the rate at which they can observe and acquire information. 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.

Giglio et al. (2021) 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 trades. We build on this literature by arguing that biases in agents' learning process, due to limited capacity of processing information, can induce correlated beliefs between stocks, which in the end induce correlated trading and interdependence between these stocks at a micro-level, and stock market comovement at a macro-level.

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

[3]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".

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. 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.

We also contribute to the extensive literature studying comovement in the stock market and interdependence between groups of stocks. There are many studies that show evidence that comovement in the stock market is not directly driven by common factors typically studied or rational behavior (Lee et al., 1991; Pindyck and Rotemberg, 1993; Froot and Dabora, 1999; Barberis and Shleifer, 2003; Barberis et al., 2005; Veldkamp, 2006; Greenwood, 2007; Kumar et al., 2016). Froot and Dabora (1999) provide evidence of stock market comovement by examining pairs of 'siamese' twin companies with different trading habitats, showing that the difference between these stock prices is related to the markets on which they are traded. Barberis et al. (2005) gives further evidence in the context of the addition of stocks to the S&P 500.

Veldkamp (2006) develops a framework in which when information is costly and (rational) investors only buy a subset of the available information, news about one asset affects the other assets' prices and asset prices comove. The mechanism we propose in which beliefs are correlated across assets is similar to the one they propose, with the difference that we do not limit the amount of information that investors consume, but we assume that the way in which they learn about sets of information is biased towards information common to more than a stock.

Barberis et al. (2005) also provides intuitions about potential mechanisms. Investors might have a 'category view' while investing, meaning that they group assets into categories and allocate funds at that level. We show that there is interdependence between stock returns and although that interdependence we attribute it to correlated beliefs, these correlated beliefs might be driven by some of the mechanisms mentioned, among many other potential mechanisms. There have been other mechanisms studied in the literature, such as comovement arising due to common analyst coverage (Muslu et al., 2014; Hameed et al., 2015; Israelsen, 2016), mutual fund ownership (Antón and Polk, 2014) and many others. We show that many of these theories can be relevant in explaining the interdependence observed across stocks, but there is still an important part explained by our informational linkages across stocks.

We also contribute to the literature studying the dynamics of optimal production and acquisition of information, which can also shed light on how investors' behavior changes when there is 'more at stake' (such as crisis periods), compared to normal periods. The main idea, deeply studied by Gorton and Ordoñez (2014); Chousakos et al. (2023); Cole et al. (Forthcoming), 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. 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 those periods.

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 Mendoza (2000); De Gregorio and Valdés (2001); Bekaert et al. (2014); Debarsy et al. (2018). Herding can be viewed as the aggregate result of the interdependence between groups of stocks. On the other hand, with our methodology, we can also study the contagion of simulated and real shocks in the stock market.

3 Empirics

In this section, we explain in more detail the intuitions behind the informational channel, how we estimate empirically its relevance, and its different impacts in periods of higher uncertainty and when considering category-learning effects.

3.1 The Informational Channel

3.1.1 Intuitions behind the Informational Channel

We believe that due to biases in the agents' learning process, such as limited attention capacity and costs to learn new information, anomalous comovements in agents' beliefs can arise[4]. These biases induce correlations of beliefs about the variables that describe the performance of stocks. We name the formation of these correlated beliefs informational linkages.

[4]We can not prove this theory with our data, so we also keep open the possibility that other channels explain these puzzle.

These links can capture many ways in which stocks can be linked with each other, some of them can not be directly observed by the data, and that is one of the novelties of our approach. Directly, these links capture common beliefs between investors, but these common beliefs may or may not be grounded in fundamentals. In Section 4.3 we show that these links are still relevant in explaining stock market comovement, even after controlling for typical observables that might link different stocks.

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 might arise.

Assume an infinite number of periods and a representative agent that 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 pair-common factor between stocks.

These factors are identically and independently distributed across periods 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 again is unknown but inferred by the agent.

The agent learning process is determined by how they interpret the signal regarding each factor. The agent has two learning biases, the first arises because they over-weight the precision of the common factor signal. This assumption is based on the learning costs of information about each individual stock are usually higher than the costs of learning about a common set of the market.

Motivated by this intuition, agents might wrongly believe that the signal regarding the common factor is more informative that it is in reality, over-weighting the precision of this signal.

The second bias in the learning process is related to agents over-weighting the pair-common factor signal. The intuition behind this assumption is that individuals have limited attention capacity and thus learn in terms of groups/categories (Barberis and Shleifer, 2003; Barberis et al., 2005). Note that this category-learning channel can be generated even if the costs of learning about the stock-specific factor are zero.

These biases in the learning process will generate correlated agents' beliefs across stocks, which will create interdependence in idiosyncratic stock returns.

In a nutshell, agents' learning process and their beliefs will determine stock returns, since investors' trading drives returns of stocks. Agents form their beliefs and trade with the information they receive. Following Bernales et al. (2024), agents receive signals of information on idiosyncratic factors and signals on common factors. The correlation in agents' beliefs is induced by biases in the agents' learning process: agents overweight the precision of new information related to common factors, this induces correlated errors in agents' beliefs, which generates interdependence between stock returns. This is what we call the informational channel.

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

3.1.2 Quantifying Informational Linkages

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

To proxy agents' belief errors, we use monthly analyst forecast errors on the Earnings Per Share (EPS)[5] of each stock. Then we calculate the correlations of these forecast errors. 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. To avoid results being driven by a small number of analysts, we require a minimum of five analysts to calculate $NRMSE_{i,t}^{EPS}$.

The forecast errors observed in the previous step might be driven by common factors, such as a common shock to the economy, or a specific group of it, that made forecasts mistaken for all stocks. We follow

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

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 (over 0.5), we de-factor the measure until this is weakly CSD (below 0.5)[6]. 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. Our baseline model uses one common factor. But we have tested with variations of this number. 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.

I 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, I follow Bailey et al. (2016) and apply a multiple testing procedure (Holm, 1979):

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

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 of this is that an element w_{ij}^{I} of the matrix of informational linkages $\left(W^I\right)$ 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. I allow for different degrees of intensity of the link based on $\hat{\epsilon}_{i,j}$, multiplying w_{ij}^{I} by $\hat{\epsilon}_{i,j}$ and row-normalizing. 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 an $N \times N$ matrix of informational linkages that will have the following

[6] If the degree of CSD is below 0.5, Bailey et al. (2016) 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.

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).

Figure 2 illustrates the informational linkage matrix that we estimated. Connections between stocks are in white, while sparse elements (stocks not connected) are in black. Firms are divided into those that have been part of the S&P 500 index, and those that have not been part. The red line separates these groups[7]. We estimate a different informational linkage matrix by year, to capture the dynamic component of these connections. Figure 2 (a) includes our estimated matrix of informational connections for 2005, while Figure 2 (b) for 2008. It can be seen that in 2008 there are more sparse elements. The changes in these connections are discussed in Section 4.3.

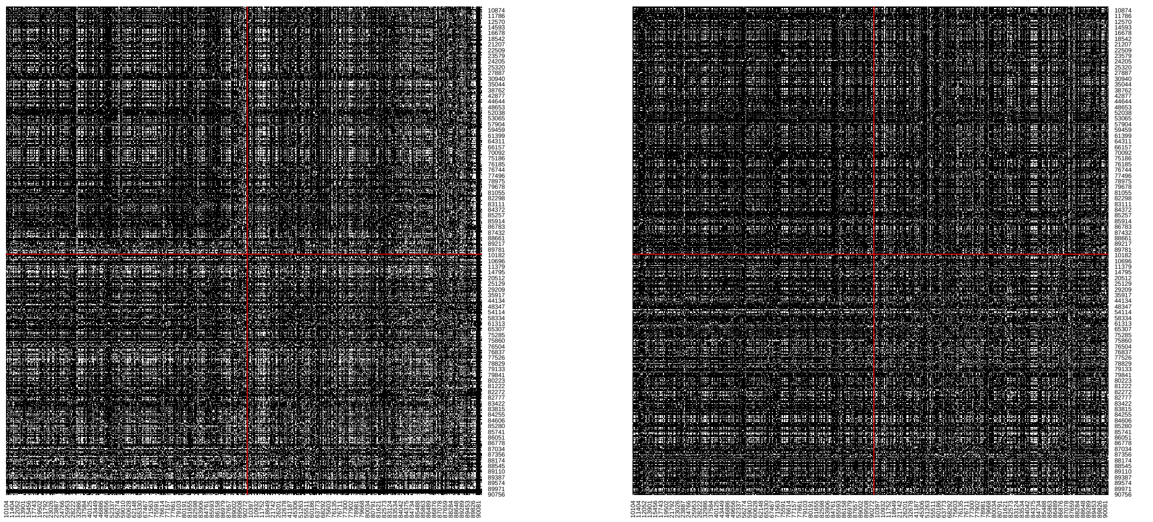
These linkages are based on correlated beliefs, but they are not necessarily irrational. They might be capturing linkages between stocks that are not directly observable by data, they might be capturing an input-output linkages structure, between many other possible structures. We do not take a stand on that, but we do assess concerns about our linkages being driven by observable categories in which investors might group stocks.

For instance, the reader might think that our linkages are just capturing linkages between stocks in the same industry, stocks that are part of an index (such as the S&P 500 index), stocks that are geographically close to each other, or stocks that are connected by common analyst coverage. We give higher detail about

[7] The groups of each quadrant (separated by red lines) are:

$$W^I = \begin{bmatrix} W^I_{SP,SP} & W^I_{SP,NSP} \\ W^I_{NSP,NSP} & W^I_{NSP,NSP} \end{bmatrix}$$

Figure 2: Matrix of Informational Linkages



(a) W^I_{2005} (b) W^I_{2008}

Notes: This figure shows the matrix of informational linkages for (a) 2005 and (b) 2008. Each element of the matrix indicates whether a pair of stocks is connected (white) or not connected (black). The red lines separates stocks that have been part of the S&P 500 (left to the red line) with stocks that have not been part (right). More details in the text. Source: Own elaboration based on I/B/E/S.

this in Section 4.3.

3.2 Empirical Implementation (S2SLS) and Data

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 comovement. Since we have a high dimensional matrix of informational linkages, Spatial Econometric models are of particular use.

The main objective of these models is to accommodate cross-sectional dependencies among observations and to model in an explicit way interactions between observations. In our context, it allows us to explicitly account for a matrix of interdependences between stocks by including an interaction matrix (our informational linkages matrix).

This approach also allows us to include a single or multiple sources of interdependences with one or several interaction matrices, 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).

Each value in the informational linkage matrix indicates if a pair of stocks is linked. If they are, these two securities are interconnected, if one of them suffers a shock it will transmit to the other. These interconnections give origin to some unexplained idiosyncratic stock comovement, which motivates this paper.

As expressed in the Introduction, we recognize that fundamentals or common factors can explain a part of stock market comovement. We try to understand the component of stock returns that these factors should not explain, the idiosyncratic component. 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, with the following steps:

1. Compute the excess return of each asset by subtracting the effective daily return of the asset by a risk-free instrument at a daily basis.

2. At a security level, keep idiosyncratic returns by de-factoring the assets excess returns with Fama-French's factor.

$$\mathbf{er}_t = \boldsymbol{\alpha} + \boldsymbol{\beta}\mathbf{F}_t + \mathbf{u}_t \tag{4}$$

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) [8]. Note that α is a vector of $N \times 1$ stock-specific slopes and β is a matrix of $N \times 6$ coefficients for each firm-factor.

[8](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

We keep the residuals, $\hat{\mathbf{u}}_t$, from the OLS estimation of Equation (4), which we call the 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.

Theoretically, idiosyncratic returns should not comove. In Figure 1 we show that they do comove and this can be explained by informational linkages.

Having described how we estimate idiosyncratic returns, we move into the model with which we understand the relevance of informational linkages in describing the interdependence of "idiosyncratic" stock returns. Consider N firms over T periods. Denote by $\mathbf{r}_t$ the vector of idiosyncratic returns at period $t \in T$ for a set of stocks (i). The model to be estimated is[9]:

$$\mathbf{r}_t = \rho \mathbf{W}_{t-1}^{I} \mathbf{r}_t + \mathbf{X}_{t-1}\beta_i + \boldsymbol{\mu} + \gamma_t + \mathbf{v}_t \tag{5}$$

Where $\boldsymbol{\mu}, \gamma_t$ are stock and time fixed effects, $\mathbf{W}_{t-1}^{I}$ is our $N \times N$ matrix of informational linkages (previously estimated) that varies each year, $\mathbf{X}_{t-1}$ 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 W_{t-1}^{I} are zero. Thus, the idiosyncratic return of a stock i is not used as an explanatory variable for the same stock.

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 transmission of shocks across securities.

Another concern that might arise is that correlated beliefs about a pair of stocks could arise simultaneously

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

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

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

with what determines idiosyncratic returns. Note that W^I_{t-1} is lagged one year with respect to $\mathbf{r}_t$, we do this to address this concern. We also construct 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 W^I that is fixed across time.

Idiosyncratic stock returns might 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 [10]. Note that in our specification, the idiosyncratic returns of stocks $i \neq j$ are used (weighted) to explain the idiosyncratic return of stock i. This might 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 the appendix, you can find details of this estimation method. The first stage is:

$$W^I_{t-1}\mathbf{r}_t = \theta \mathbf{H_n} + \mathbf{u}_t \tag{6}$$

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 $\widehat{\theta} = \begin{bmatrix} \widehat{\rho} \\ \widehat{\beta}_i \end{bmatrix}$. Then, we use $\widehat{W^I_{t-1}\mathbf{r}_t}$ in the second stage as in Equation (5).

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 got $N > T$. Moreover, in GVARs, there is usually no distinction between different types of variables in the sense of them being all

[10]Since $W^I_{t-1}\mathbf{r}_t$ is a regressor, that source of endogeneity is important for us.

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 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.

In terms of data, we use data of EPS analyst forecasts from I/B/E/S, financial ratios from Compustat, and security returns from CRSP. We estimate the spatial model with data from January 2005 until December 2019. However, to construct each 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, W^I_{2005} uses data from 2000 until 2004, 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[11].

We would like to have as many stocks as possible, but the core of our data comes from three different data sources, which do not always have a great match. Due to the methodology we are using, we require securities to be present in most of the timespan of our estimation. Considering these restrictions, we can keep a subset of 771 securities. About half of them have been at some point part of the Standard & Poor's Index, and half of them have never been part of it.

We recognize that by keeping stocks that have forecasts by analysts and that are present in a big number of years, this might not be representative of the whole stock market. The distribution of our subset of securities, as compared to a bigger set of securities, is somehow bigger and more resilient. Again, this might raise some concerns regarding the results that we will show. We believe that if our securities are bigger and have been present for more time, it is easier to learn about them. This implies that the results that we will

[11] 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 might arise due to this. We acknowledge that this assumption might 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.

show are in some sense a lower bound.

Table 1 shows a summary statistics of our subset of securities[12].

Table 1: Summary Statistics

	Min.	p25	p50	p75	Max.	Mean	St.Dev.	Obs.
Idios. Return	-18.53%	-4.24%	-0.22%	3.83%	20.46%	-0.04%	8.76%	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 estimation being driven by outliers.

4 Results

In this section, we aim to learn if the proposed informational channel of interdependence 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 comovement arises. In Section 4.1 we analyze the relevance of the informational channel, while in Section 4.2 we analyze the difference in the strength of the informational channel on periods of higher uncertainty and when considering category-learning effects in Section 4.3. In Section 4.4 we analyze the robustness of our results.

4.1 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. To be sure that results

[12]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.

are not driven by the control variables selected, we estimate our baseline model described in Equation 5 with different sets of controls, it can be observed that the coefficient does not vary depending on the set of controls used. This table tells us that a negative shock in a security i does not only impact security i but also all of those securities $i \neq j$ linked in an informational way to security i.

This is strong evidence that the informational channel can explain idiosyncratic stock returns, hence, it is also able to explain stock market comovement. It is relevant to mention that the interpretation of the coefficient is not marginal. Intuitively, the coefficient is a measure of the strength of the informational linkages matrix. Note that Equation (5) can be written as:

$$\mathbf{r}_t = (I_N - \rho \mathbf{W}^I_{t-1})^{-1}(\mathbf{X}_{t-1}\beta_i + \boldsymbol{\mu} + \gamma_t + \mathbf{v}_t) \tag{7}$$

Table 2

	Dep. Var.: Monthly Idios. Return		
	(i) Random Controls	(ii) Financial Ratios	(iii) Full Controls
W	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 5 randomly normally distributed controls. The second includes as controls: 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.

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

4.2 Uncertainty Periods

In this section, we analyze how important the informational channel is in periods with higher uncertainty. The question that motivates this analysis is if learning biases change in periods with higher uncertainty. The learning process of individuals might change when they face more stress. There are two strands of the literature that implicitly study this idea. The first is related to the incentives of producing and acquiring information. Gorton and Ordoñez (2014) have studied this in detail. They propose a model where firms have debt that is backed by collateral. Information production about the backing collateral is costly to produce, and agents do not find it optimal to produce (costly) information at every date, which leads to a depreciation of information over time in the economy.

The key dynamic in their model concerns how the perceived quality of collateral evolves if (costly) information is not produced.

After a credit boom, in which more and more firms borrow with debt backed by collateral of unknown type (but with high perceived quality), a negative aggregate shock affects a larger fraction of collateral than the same aggregate shock would affect when the credit boom was shorter or if the value of collateral was known. A negative aggregate shock may or may not trigger information production, because investors have more to lose when they do not have that information. There may be no effect. It depends on the length of the credit boom.

Cole et al. (Forthcoming) studies a similar question but in the context of investors buying government debt. In their model, governments finance their budgets by selling bonds in a sequence of auctions. This leads to information rents for investors who know more about the fundamental values of bonds. Information is particularly valuable during periods of heightened uncertainty in which default risk can vary substantially from auction to auction, such as when there are concerns about a country's solvency or policy stance.

In such circumstances, some investors may start acquiring information before bidding for bonds, chasing away other investors who do not become informed and instead move more of their wealth to other countries.

As a result of these interactions between information acquisition and cross-country flows, fundamental shocks in one country may affect yields and portfolio choices in other countries even when there are no fundamental linkages between them. They develop a model that shows that these interactions are rich and account for many features of the 2010 Eurozone sovereign debt crisis.

Applying these studies to our context, both of these papers suggest that the acquisition and production of information might change during periods of higher uncertainty. Thus, in these periods investors would demand and be able to access information in a different way. Thus, their learning biases might change during these periods. Particularly, these studies suggest that the intensity of informational linkages (ρ in our baseline specification) should decrease during these periods.

The second strand of the literature is related to herding and contagion theories. This has been deeply studied, as in Kaminsky and Schmukler (1999); Calvo and Mendoza (2000); De Gregorio and Valdés (2001); Forbes and Rigobon (2001); Basu (2002) among many other contributions in the literature. The main intuition of this theory is that investors tend to follow the "market" instead of making their own assessments of fundamentals, since market portfolios embody relevant information[13]. Investors might want to follow the market because the cost of acquiring information can be higher than the potential gains over market portfolios.

Calvo and Mendoza (2000) also shows that investors have fewer incentives to gather good news in a bullish market, while incentives to verify good news are increasing in a bearish market. Hence, good rumors generate more herding behavior in good times than in bad times. Thus, this paper also suggests that the intensity through which shocks propagate through the informational channel is lower in periods of higher uncertainty.

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.

We study how learning biases change during periods of higher uncertainty by adding a new matrix of informational interdependence that interacts with the original W_t^I with a dummy for periods of higher uncertainty and estimating its relevance with the following specification:

$$\mathbf{r}_t = \rho \mathbf{W}_{t-1}^I \mathbf{r}_t + \rho_{Uncert.} \mathbf{W}_{t-1}^I \mathbf{Uncert}_t \mathbf{r}_t + \mathbf{X}_{t-1}\beta_i + \boldsymbol{\mu} + \gamma_t + \mathbf{v}_t \tag{8}$$

[13]The intuition of investors following market portfolios should not be of great concern in our study, since we are using idiosyncratic returns. The de-factoring process we apply to get idiosyncratic returns considers a factor of the market portfolio. Now, we still consider this literature relevant for our context since it can be applied to any other group. For instance, herding might be at an industry level.

Where $\rho_{Uncert.}$ is the new coefficient to be estimated and $Uncert_t$ is a dummy for periods of higher uncertainty. We estimate this specification with different measures of periods of higher uncertainty. First, we use only the periods related to the Global Financial Crisis (GFC), we followed the literature and considered it from July 2007 until March 2009. We also used the 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).

Table 3 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_t^I with periods of higher uncertainty. This means that, just as we hypothesized, there is a decrease in the intensity of the informational channel in periods of higher uncertainty. This also follows the theoretical literature we have just discussed. Intuitively, this can be due to investors knowing that in these periods they have more at stake, and focusing on learning about firms' fundamentals can have a higher reward.

Table 3: Informational Linkages and Uncertainty Periods

	Dep. Var.: Monthly Idios. Return							
	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	138780	138780	138780	138780	138780	138780	138780	138780
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.

It is interesting to notice that the amount of connections increases in periods of higher uncertainty, seen as a decrease in the level of sparseness of W matrix, but the intensity of these connections decreases in these periods, as it can be seen in the Appendix, in Table 8.

4.3 Category Learning

Comovement in agents' beliefs (and in stock returns) might also arise due to investors learning in terms of categories. This is studied in the stylized model of Bernales et al. (2024) discussed in Section 3.1. The result from this model is based on agents wrongly believing that signals of information regarding the category-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 investors, to simplify their decisions, group assets into categories and allocate funds at the level of these categories. If there are correlated sentiments 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 comovement. even when these assets' cash flows are uncorrelated.

In this section, we study if our proposed informational channel is still relevant when considering category-learning effects. Thus, 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 4, 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.

Investors might also naturally think in terms of industries. Investors might think that news will have an impact on all securities in the same industry. We also test 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. Other categories studied in the literature are those related to geographical closeness. Investors might 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. The first is connecting firms with a threshold of 100km (5% of the distribution). If a security's headquarters is at a 100km or less distance from other securities, we say that they are geographically linked. The second one is at a state level. If two securities are in the same state, we say that they are geographically linked. To give an idea of how securities are located, Figure 4 shows the location of securities headquarters.

Finally, Israelsen (2016) studies the effect of common analyst coverage on explaining excess comovement. 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. Rational Bayesian updating leads to an increase in comovement when correlation in forecast errors increases. They show that pairs of stocks with higher forecast errors exhibit more excess comovement, even after controlling for risk.

This idea might raise some concerns about how we estimate informational linkages. We use correlated analyst forecast errors to proxy for correlated beliefs. A direct critique of our metric is that instead, it could be capturing common analyst coverage. To address this critique, we construct a matrix of common analyst coverage.

To construct this matrix, 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. We show that the intensity of the relevance of the informational channel does not vary so much when considering $W_t^{ComAnalys}$.

Table 4: Informational Linkages and Category Learning

	Dep. Var.: Monthly Idios. Return					
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.

Table 4 shows that the informational channel is still relevant even after considering each of the category-learning effects proposed. It also shows that the intensity of the informational channel increases when considering category-learning effects. 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.

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

Overall, with each of the applications shown in Tables 2, 3 and 4 it can be seen that the informational channel can explain idiosyncratic stock returns. Thus, showing interdependence between securities, which in the end generates comovement between stock returns.

4.4 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 4.1.

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 might have different levels of EPS, normalization makes this number comparable across stocks. Although after that we use the correlation of forecast errors, before that we do winsorize the level of the NMRSE at the 2% level. At this point, normalization plays a role.

We use different measures for forecast errors in Panel A 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 (iv) 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			
	(i) RECM	(ii) NSD	(iii) Fix W	(iv) $p = 0.05$
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. (i) uses the Root Mean Squared Error as a forecast error measure, (ii) uses the Normalized Simple Difference as a forecast error measure. (iii) uses a fix instead of dynamic W matrix to account for informational linkages. (iv) sets the threshold to define informational linkages strong enough to be 0.05 instead of 0.01 T-stats are below each coefficient.

5 Applications

Having shown the relevance of the informational channel, we now move into applications that we can do with our baseline model. Since we have estimated a network of linkages at a firm-to-firm level, we can easily move into 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 can generate aggregate fluctuations when network effects are relevant.

We will analyze how simulated and real-world idiosyncratic shocks propagate through the stock market. We start with a simulated one standard deviation shock to a stock 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 how well our estimated network captures indirect effects over the stock market, compared to what happened in reality.

Our baseline model allows us to identify the direct and indirect impacts of a shock over one stock. The model can be written as[14]:

$$\mathbf{r}_t = (I_N - \rho \mathbf{W}^I_{t-1})^{-1}(\mathbf{X}_{t-1}\beta_i + \boldsymbol{\mu} + \gamma_t + \mathbf{v}_t) \tag{9}$$

Then

$$\frac{\partial \mathbf{r}_t}{\partial \mathbf{v}_t} = (I_N - \rho \mathbf{W}^I_{t-1})^{-1}\mathbf{v}_t \tag{10}$$

Note that if we apply a shock in one period 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 stock "A" is connected to stock "B" and stock "B" is connected to stock "C", then a shock over stock "A" might 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) might 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

[14] 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.

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.

5.1 Simulated Shocks

In Table 6, we simulate a 1 standard deviation of idiosyncratic returns shock (on avg. 8.86%) to one firm i and estimate its indirect impacts over $j \neq i$, as in Equation (10)[15]. The first column shows that on average, a one standard deviation shock over i generates a 14.40% increase in idiosyncratic returns over $j \neq i$, relative to the mean idiosyncratic return of the cross-section of stocks. 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 does imply that the shock generates an important increase on the idiosyncratic returns of firms $j \neq i$, relative to the mean idiosyncratic return.

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 an average indirect impact of 32.65%, relative to the mean, indicating that shocks to central firms have more than twice the effects than an average firm. Column (3) focuses on the five least connected firms, provided they have at least one connection. The average indirect impact 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, and divides over all N 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 3 extends the analysis from Columns (2) and (3) and plots the average firm-level indirect effects

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

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
Impact	14.40%	32.65%	0.73%	28.54%

Notes: Estimates correspond to the average marginal firm-level indirect effects of a one standard-deviation exogenous shock to a given security i's idiosyncratic return (on avg. 8.86%) on other securities $j \neq i$, divided on the mean monthly idiosyncratic return of the cross-section of securities (0.23%). The first column applies the shock to all firms (separately), sums the indirect impact over other firms and divides it over N, the number of stocks (avg.), then it takes the average across shocked firms. 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 they 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 in 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).

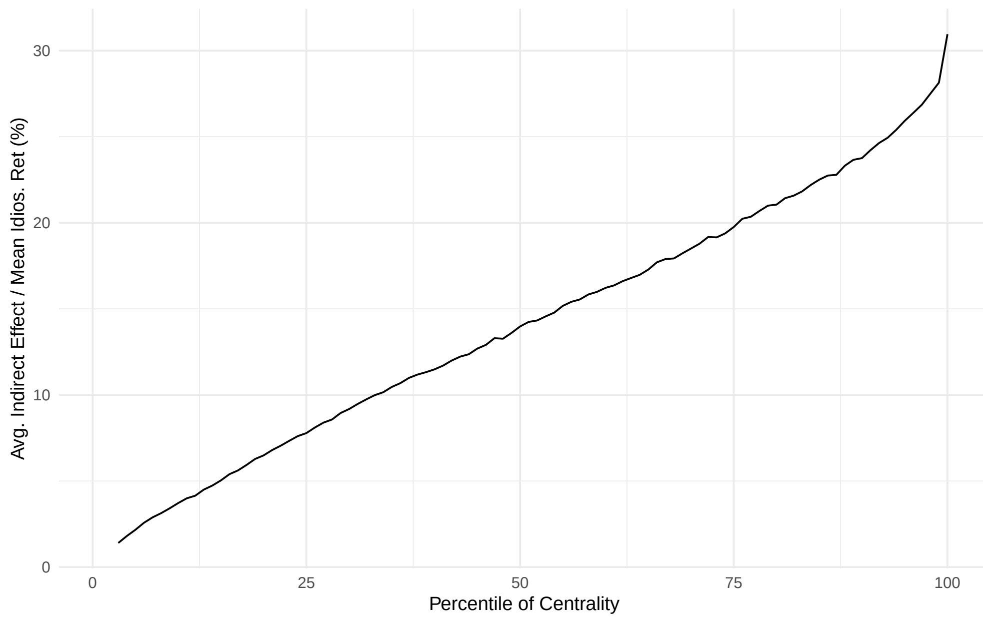
by percentile of centrality. As we have already mentioned, more central firms propagate more idiosyncratic shocks than less central firms. How relevant is this difference? In the figure it can be observed that firms with higher centrality (closer to the end of the x-axis) propagate shocks disproportionately more than firms with lower centrality, there is a clear change in slope. This means that there is a subset of firms that propagate shocks with a much higher intensity than the rest.

5.2 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

Figure 3: Indirect Impacts of a 1sd shock (7.37%) 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 sd 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).

and intense rainfall, leading to severe flooding and significant economic disruption. We will examine the stock market's response to this event.

Our approach is simple; we will analyze the stock market's response to firms affected by the shock: firms with their headquarters (HQ) in NC, SC, and VA. Then, we will 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 will 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[16], was (-0.59%/-1.55%) 37.97% of the affected stocks.

[16]This was done using W^I_{2018}.

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 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.

6 Conclusion

In this paper we show empirically that stock return comovement might arise due to analysts' learning biases, providing a new channel on the long-standing puzzle about excess stock market comovement. Learning biases might give origin to correlated beliefs between agents, which we call informational linkages. We show that informational linkages generate interdependence between stock returns at a micro-level, and stock market comovement at an aggregate level.

Our informational channel might include different channels already studied in the literature, based on the category-thinking, which suggests that investors learn and think in terms of categories, grouping stocks according to, for instance, geographical proximity, common-index (e.g. stocks belonging to the S&P 500), being part of the same industry, and stocks that share common-analysts. We show that category thinking is relevant and explains the interdependence between stock returns, but our informational channel still stands while considering these theories.

We also study the dynamics of these potential learning biases, showing that in periods of higher uncertainty, learning biases are lower relative to normal periods. This means that the intensity of interdependence between stock returns due to our informational linkages diminished in periods of higher uncertainty, this might be due to investors' knowing that in periods of higher uncertainty, they have more at stake and higher incentives to inform themselves better.

Finally, we use our estimated matrix of informational linkages and apply simulated and real-world shocks. With our simulated shocks, we find that idiosyncratic shocks to one firm have an important propagation through the stock market over other stocks (indirect effects), with more central firms propagating the shocks disproportionately more. With our real-world shock, we show that our estimated informational linkage matrix can capture the real-world propagation of shocks.

Our channel is based on biases in the learning process of investors. If investors understand their biases, they might 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 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, future research could explore this area with more detail. In particular, it would be an interesting exercise to compare informational linkages to other sources of linkages between stocks, such as input-output linkages, geographical connections, among others. Further research could also explore how relevant is this channel in a different context, such as emerging countries' stock markets.

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. Forthcoming. "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.

De Gregorio, José, and Rodrigo O. Valdés. 2001. "Crisis Transmission: Evidence from the Debt, Tequila, and Asian Flu Crises." *The World Bank Economic Review* 15 (2): 289–314, http://www.jstor.org/stable/3990267.

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, K.A., and E. Dabora. 1999. "How Are Stock Prices Affected by the Location of Trade?" *Journal of Financial Economics* 53 (2): 189–216.

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, 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.

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

Appendix A: Spatial Two-Stage Least Squares (S2SLS) Details

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

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

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

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

Figure 4: Securities Headquarters

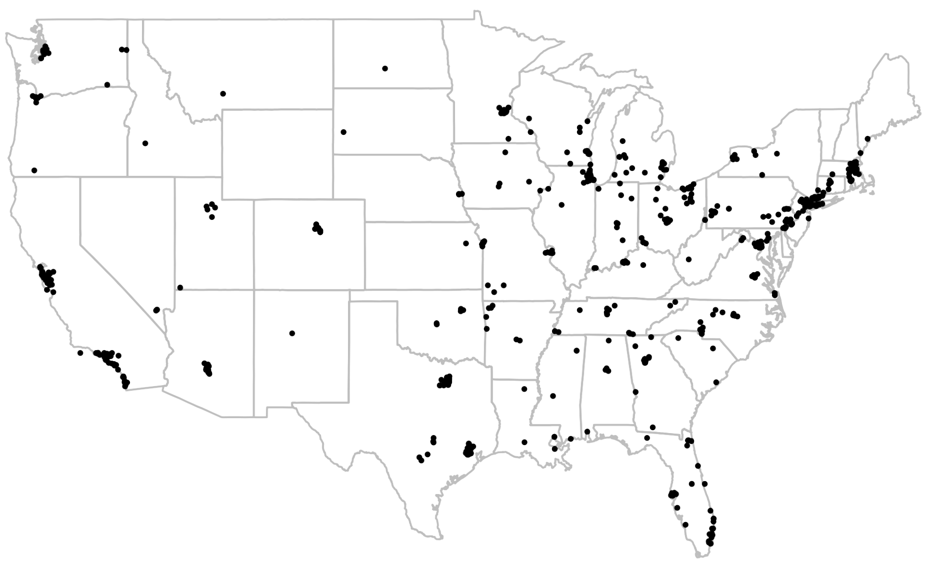


Table 8: Level of Sparsness and Intensity of W by year

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%