UNIVERSITÀ COMMERCIALE “LUIGI BOCCONI”

PhD SCHOOL

PhD program in Economics and Finance
Cycle: 32
Disciplinary Field: SECS-P / 11

Essays on Information in Finance

Advisor: Mariano Massimiliano CROCE
Co-Advisor: Miles Bernard GIETZMANN

PhD Thesis by
Isabella WOLFSKEIL
ID number: 3025407

Year 2022

Table of Contents

Summary

This thesis is a compilation of three separate papers. In the first paper I investigate the link between firms' voluntary disclosure strategies on social media and their equity returns. I construct a novel and comprehensive database of over 7 million tweets posted by S&P 1500 firms and use text analysis methods to assess the effect of corporate tweets on announcement returns. I find evidence consistent with firms using the timing, tone, and content of tweets strategically. Firms with negative earnings surprises have higher announcement returns when they tweet about financial news, suggesting that firms can use social media to bolster their stock prices during periods of poor performance. This result holds mainly for firms with higher retail investor ownership, consistent with social media being a primary information source for investors with a high cost of information acquisition and processing.

The second paper is a joint work with M. J. Arteaga-Garavito, M. M. Croce, and P. Farroni. We quantify the exposure of major financial markets to news shocks about global contagion risk accounting for local epidemic conditions. For a wide cross section of countries, we construct a novel data set comprising (i) announcements related to COVID19, and (ii) high-frequency data on epidemic news diffused through Twitter. Across several classes of financial assets, we provide novel empirical evidence about financial dynamics (i) around epidemic announcements, (ii) at a daily frequency, and (iii) at an intra-daily frequency. Formal estimations based on both contagion data and social media activity about COVID19 confirm that the market price of contagion risk is very significant. We conclude that prudential policies aimed at mitigating either global contagion or local diffusion may be extremely valuable.

The third paper is a joint work with Lucia Alessi, Brunella Bruno, Elena Carletti and Katja Neugebauer. We analyze the determinants of coverage ratios and their components (NPLs and loss loan reserves) in a large sample of European banks. We find that bank-specific factors, and in particular credit risk variables including forward-looking indicators, matter the most. We also uncover that coverage ratios do not adjust sufficiently when asset quality deteriorates but that high-NPL banks tend to be relatively better covered. At the country level, specific macroprudential levers as well as developing NPL secondary markets enhance bank coverage policy. Our findings emphasize the importance of micro prudential oversight and call for more stringent macro policies in high-NPL countries.

I would like to express my sincere gratitude to my advisors Prof. Mariano Massimiliano Croce and Prof. Miles Gietzmann for the invaluable advice, continuous support, and patience during my PhD study. I could not have imagined having better advisors or mentors. Besides my advisors, I would like to thank the rest of my thesis committee: Prof. Elena Carletti and Prof. Alberto Manconi, for their insightful comments and encouragement. My sincere thanks also goes to the entire Finance Department at Bocconi for guiding me with their immense knowledge and for the many hard question which helped shape my research. Any remaining errors are my own.

Chapter 1. Tweeting in the Dark: Corporate Communication and Information Diffusion

1. Introduction

Social media has changed the way firms communicate with investors by giving them a direct, instantaneous, and network-enhanced communication channel. Firms can now directly transmit information to shareholders through Twitter, Facebook, YouTube, and Instagram, among others. In 2013, the SEC announced that companies could use social media to disseminate material information as long as investors were alerted that social media was being used to announce such information. Despite the regulatory attention social media has received, these channels remain voluntary forms of communication. This means that managers have the discretion to disclose or withhold information on social media as they see fit. Given the current regulation, corporate disclosures on social media must be studied in a setting in which firms optimally choose their disclosure strategy and investors anticipate that firms may disclose news strategically.

Managers' incentives are likely to be an essential determinant in the information disclosures investors observe on social media. In fact, recent empirical evidence shows that firms are more likely to disclose good news than bad news on their social media platforms (see Jung et al. (2018)). Yet, the effects of strategic disclosure and stock prices remain understudied. In light of these facts, this paper provides a novel empirical investigation to address the following research question: what is the link between firms' disclosure strategies on social media and their equity returns?

I exploit firms' discretionary use of social media in disseminating quarterly earnings announcements to examine the relationship between disclosure strategies and equity returns at daily and intradaily frequencies. By focusing on the voluntary disclosure of information on Twitter following mandatory earnings announcement events, it is possible to disentangle the effect of the voluntary disclosure decision from the effect of the news itself. The SEC requires that firms announce their earnings results at the end of each fiscal quarter. The market actively anticipates these announcements and any deviation from the market's expectation ultimately determines the reaction of the stock price to the announcement. Using the deviation from analysts' forecasts, i.e., the actual earnings per share minus the analysts' forecast, I can control for the news itself and isolate the impact of the voluntary disclosure decision on equity returns.

My study focuses on the popular social media site Twitter, which was created in 2006 as a free service that allows users to communicate through short messages of up to 280 characters, known as "tweets." I focus on Twitter because, unlike many other social media platforms, Twitter was designed for sharing news and information in real time. Also, it has surpassed other social media platforms in terms of general corporate adoption and for disseminating investor-related announcements (see Jung et al. (2018). I construct a novel and comprehensive dataset that aggregates over 7 million individual tweets and represents the complete tweeting history of more than $1,000$ firms from January 2014 through December 2018. One of the primary challenges underlying the research design is the detection of financial news disclosure on Twitter. I use text analysis methods to identify tweets related to earnings announcement news and focus on the tweets over a short window around the announcement.

In the empirical part of my study, I document three important results. First, I find that tweeting has an asymmetric effect on announcement returns, depending on whether firms tweet about financial news on positive or negative earnings surprise days. In particular, firms with negative earnings announcements have higher announcement returns when they tweet about their earnings news. A separate high-frequency analysis supports this result. The speed of information flow on Twitter creates a unique setting to study investors' immediate reactions to tweets about financial news. I find that when firms with negative earnings surprises tweet about financial news, their abnormal cumulative returns appreciate substantially in the 30 minutes following the tweet. Second, I provide evidence that the dissemination of public information on social media matters more for retail investors, which tend to have higher information acquisition and processing costs. Finally, I employ natural language processing techniques to investigate the strategic use of tone and information content in tweets.

The results I document are consistent with firms using Twitter strategically. In line with previous research, I find that firms are more likely to tweet about their financial results in instances in which they meet or beat analysts' estimates. Next, I study the tone of financial news disclosures on Twitter. Generally, tweets have a positive linguistic tone, independent of whether firms disclose financial news around a positive earnings surprise announcement (good news) or a negative surprise earnings announcement (bad news). Finally, I compare the information content of tweets around good news and bad news. Notably, tweets are less likely to mention "earnings per share" and more likely to mention "dividends" on days with negative earnings surprises. Both earnings per share and dividends are closely watched by investors and communicate the financial well-being of a firm. These results suggest that firms not only strategically choose when to tweet about earnings announcements but also what kind of information to include in their tweets.

I use a model to shed light on the mechanisms through which strategic voluntary disclosure impacts investors' expectations and, ultimately, the price of firms' equity. I examine the effects of strategic voluntary disclosures made after earnings announcement events using a framework introduced by Goto et al. (2009). The model analyzes disclosures in terms of a verifiable reports framework to capture the broad limits imposed by the accounting system. Even though managers have discretion in disclosing information on social media, corporate disclosures must be truthful. In this model, a firm has multiple projects, each which can succeed or fail. The firm's manager observes some of these outcomes, while investors observe only the public disclosure made by the manager. The manager is free to disclose some or all of what he knows at an interim date, though they cannot concoct false information. The disclosure policy of the manager is driven by the objective of maximizing the price of the firm at each date. At the same time, investors appropriately anticipate the manager's disclosure policy and price the firm accordingly. This gives rise to a game of incomplete information.

I augment Goto et al.'s (2009) model to include a mandatory disclosure event that occurs at the start of the game. Each firm announces its quarterly earnings results to the market, and investors update their prior probability that a dimension of the firm will be successful. If a firm reports earnings below (above) the market's expectations, the expected probability that a business dimension succeeds becomes lower (higher). I interpret the positive (negative) tone of financial news tweets as a disclosure of a success (failure). In particular, I examine two manager strategies: one in which the manager follows a strategic disclosure policy (only disclosing successes), and another in which the manager follows a full disclosure policy. The model shows that firms with a negative earnings surprise have higher expected returns when moving from a full disclosure policy to a strategic disclosure policy. The intuition is that the marginal benefit of strategically disseminating information on social media is higher for firms that are less likely to have good news to disclose. Hence, the model predicts that stock prices will rise more for firms that follow a strategic disclosure policy following a relatively poor earnings announcement.

A key assumption in my theoretical framework is that investors are uncertain about the information endowment of managers. The probability that a manager is informed about the outcome of a business dimension at the interim date of the model captures the relative level of information asymmetry. This parameter can be thought of as the level of investor sophistication. Hence the more information asymmetry there is, the less sophisticated the investors tend to be. The model predicts that the jump in expected returns, when going from a full disclosure policy to a strategic one, increases with the relative level of information asymmetry.

Given the short period between financial news tweets and mandatory quarterly earnings announcements, often just a few hours, it is reasonable to assume that investors with a high cost of information processing may be uncertain about the information endowment of managers when they read a financial news disclosure on Twitter. Inattention may seem unwise; however, if time and attention are costly, such behavior may be entirely rational (see, for example, Hirshleifer and Teoh (2003)). In general retail investors have a higher cost of information acquisition and processing, and therefore the marginal benefit of strategically disseminating information on social media is higher for firms with more retail investor ownership. In line with this prediction, I find that the positive relationship between tweeting after a negative earnings announcement and daily returns is stronger in firms with relatively high retail ownership. Moreover, in a separate analysis looking at investors' demand for information, I find that tweeting about financial news is associated with higher demand for SEC filings. This result further supports the hypothesis that investors who rely on Twitter for information suffer from limited attention biases.

This study contributes to three strands of research, of which the first concerns investor attention and asset prices. Previous work has focused primarily on modeling and empirically documenting the effects of investors' limited attention. In this literature limited attention is used to help explain pricing phenomena such as predictable price moves (Cohen and Frazzini, 2008), post-earnings announcement drift (DellaVigna and Pollet, 2009), under- and overreactions to news (Hong and Stein, 1999), and return comovements (Peng and Xiong, 2006). In these studies, firms do not actively take advantage of investors' attention; in contrast, my study shows that firms exploit investors' limited attention to support their price, especially when the firm is performing poorly. Social media gives firms more control over their information environment. In the case of Twitter, the 280-character limit allows firms to select certain information from an announcement which investors will read first. This is especially important since individuals have the tendency to attend less to information that requires greater cognitive processing, and therefore the short format of tweets can increase the salience of selected information.

This study also contributes to the literature studying how media and stock prices. Huberman and Regev (2001) was one of the first papers to establish that newspaper articles can affect stock returns, even in the absence of new fundamental information. Fang and Peress (2009) and Fedyk (2018) show that the effects of media on asset prices, in the absence of new information, may be driven by the role media plays in alleviating informational frictions. My findings indicate that the results of this literature are also true for new types of media, such as Twitter. Furthermore extant research studies media produced by third-parties (Engelberg and Parsons (2011), Tetlock (2007)), by contrast I study firm-initiated media.

Finally, I contribute to the new literature evaluating the role social media plays in financial markets. Bartov et al. (2017) investigate individual investors' use of social media to share information and insights about stocks, and they show that the aggregate opinion from these tweets can predict a firm's future quarterly earnings and announcement returns. Blankespoor et al. (2014) examine how the use of social media by tech firms is associated with improved market liquidity. They find that additional dissemination of firm-initiated news via Twitter is associated with lower abnormal bid-ask spreads and greater abnormal trading depths. Bhagwat and Burch (2016) investigate whether Twitter provides firms an effective and strategic way to mitigate investors' limited attention and find that when a firm's earnings surprise is small and positive, the magnitude of announcement returns is higher. Finally, Jung et al. (2018) study whether firms use social media to strategically disseminate financial information and find that firms are less likely to share news via Twitter when the news is bad and when the magnitude of the bad news is worse, consistent with strategic behavior. I complement this literature by studying the effects of strategic discloses through Twitter on asset prices. To the best of my knowledge, this paper provides the first empirical evidence on the strategic information content and tone of tweets across positive and negative earnings surprise days. This paper provides the first high-frequency analysis of returns around individual corporate tweets– by focusing on a short time frame of just 30 minutes before and after each tweet, this analysis helps alleviate the concern that results are driven by something other than the firm's tweeting activity.

The remainder of the paper is organized as follows. Section 2 describes the dataset and section 3 discusses the regulatory setting of disclosures using social media. In section 4 I introduce the theoretical background and empirical implications, and in section 5 I detail the empirical methodology and results. Section 6 presents robustness analyses and section 7 concludes.

2. Institutional Background and Data

A. SEC rules on social media

The SEC has embraced social media and other information technologies in an effort to promote widespread access to corporate information (SEC, 2013). In 2013, the SEC officially stated that social media could be used as a channel for the disclosure of material nonpublic information and provided guidance on the application of Regulation Fair Disclosure (Reg. FD) to social media (SEC, 2013).[1] Nevertheless, social media remains generally unregulated. More specifically, firms are not prohibited from increasing the dissemination of good news and minimizing the dissemination of bad news on social media.

In this paper, I investigate the use of social media to disclose earnings announcement news. Because of the careful regulation around earnings announcements, it is likely that firms will only disclose earnings news on social media if an official disclosure accompanies this disclosure to the SEC. The SEC requires most listed companies to file a Form 10-Q (quarterly financial report) within 40 days of the end of the quarter.[2]. In the days leading up to the earnings announcement, firms can discuss their preliminary earnings results on social media as long as the firm files a Form 8-K (current report), notifying the SEC and market participants of the impending information disclosure.[3] Due to the importance of information released during earnings announcements, communication of earnings news is carefully regulated. Therefore, it is reasonable to assume that messages on Twitter serve to

[1]On July 3, 2012, the CEO of Netflix, Reed Hastings, posted the following message to his personal Facebook page: "Congrats to Ted Sarados, and his amazing content licensing team. Netflix monthly viewing exceeded 1 billion hours for the first time ever in June." The nonpublic information disclosed in the tweet, 1 billion hours, represented a 50% increase in viewing hours from Netflix's January 25, 2012, announcement. Netflix's stock price rose from $70.45 at the time of Hastings's Facebook post to $81.72 at the close of the following trading day. Because material and nonpublic information was exclusively disclosed through Facebook and Netflix had not previously informed shareholders that the CEO's Facebook page would be used to disclose nonpublic information, Hastings's post was found in violation of Reg. FD.

[2]Nonaccelerated filers with a public float of less than $75 million are granted 45 days. Companies typically file this report and their Form 10-K (annual financial report) in the last two days of the required filing period (Amir and Livnat, 2005)

[3]It is common practice for firms to disclose preliminary earnings results. Amir and Livnat (2005) find that 80% of firms in their sample consistently issue preliminary earnings announcements—on average, 26 days after quarter-end.

broaden the dissemination of announcement information or highlight specific aspects of an earnings announcement rather than reveal new information.

Prior studies investigate the information content, timing, and tone of financial statement disclosures (Rogers et al., 2011; Kothari et al., 2009; Davis et al., 2015). This study shows that in addition to the impact of the disclosure itself, the distillation and dissemination of financial disclosures can affect how investors process information.

B. Why Twitter?

The general goal of this paper is to examine the role of social media in the disclosure of corporate information. However, from a practical point of view, there are many reasons to focus on the Twitter platform. Twitter, a micro-blogging network intended for sharing news, content, and information, is the social media platform most widely adopted by S&P 1500 firms (Jung et al., 2018). Twitter connects more than 300 million monthly active users who post, read, and interact with short messages known as "tweets". Unlike many other social media platforms, Twitter has a strong emphasis on real-time information–this enables firms to broadcast financial news directly and instantaneously to a large social network. Increasingly, investor relations departments are using Twitter to reach investors with messages about earnings announcements, management changes, and public relations crises. A growing number of companies are even beginning to create Twitter accounts specifically for investors, for example, Ford Motor Co. (@FordIR), T-Mobile (@TMobileIR), and CVS Health Corp (@CVShealthIR).

Given that investors' information processing capacity is not infinite, there are a number of reasons Twitter may be a primary source of information for some investors. First, standard asset-pricing models typically assume that markets distill new information and incorporate it into their expectations instantaneously–in reality, such distillation and estimation is limited by investors' cost of acquiring and processing information (see Cohen and Frazzini (2008), DellaVigna and Pollet (2009), Hirshleifer and Teoh (2003), Hong and Stein (1999), Peng and Xiong (2006)). The 280-character limit on tweets, the equivalent on average 45 words, can potentially increase the salience of the information. Salience determines which information will most likely grab people's attention and have the greatest influence on their perception of the world. Second, unlike many other important information channels such as the business press, analysts' reports, and newswire services, Twitter is free, reducing the upfront costs of acquiring corporate information. Finally, Twitter is a push technology, and therefore, firms can initiate the information transaction rather than wait for investors to request the information. Consequently, potential investors who might not otherwise seek out information can have it at their fingertips.

C. Data collection and sample selection

To study how disclosure strategies shape the link between corporate information dissemination on Twitter and stock returns, I construct a dataset of 7,132,461 individual tweets posted by S&P 1500 firms from January 2014 through December 2017. This firm-tweet data is merged with financial data and market data to relate tweeting activity to announcement returns and short-run continuations in returns.

I begin with an initial sample of 2,454 firms, which includes all historical S&P 1500 index constituents from 2006 (the year Twitter was founded) through 2017. From the starting sample of 2,454 firms, I identify 1,215 firms with active Twitter accounts.[4]

In the Appendix I document that, on average larger firms and, incremental to size, firms belonging to the S&P 500 index have a higher probability of having a Twitter account. This result suggests that having a Twitter account is not a substitute for overall visibility but rather a complement to it. Firms with lower book-to-market ratios and firms with relatively higher valuations than their industry peers have a higher probability of having a Twitter account. Technology companies and other companies in industries that have fewer physical assets tend to have low book-to-market

[4] I started the search on each firm's corporate website. If no Twitter handle was mentioned on the corporate website, I proceeded to search directly on Twitter. The search was conducted in October 2017; therefore, the sample is composed of firms that had active Twitter accounts in October 2017.

ratios and thus are more likely to have Twitter accounts. In addition, firms in more innovative, knowledge-intensive industries also tend to have a higher probability of having a Twitter account. Please refer Appendix Table A1 for details.

After gathering the sample of Twitter usernames, I assemble a complete history of tweets generated by the 1,215 accounts from January 1, 2014, through December 30, 2017, resulting in a sample of 7,132,461 individual tweets. To isolate firm-initiated content that is visible to the firms' followers, I exclude tweets that are reply tweets and retweets.[5] This process reduces the sample to 3,305,257 individual tweets.

Quarterly earnings announcement dates and analyst consensus forecasts are obtained from Compustat and I/B/E/S, respectively. Daily stock prices are obtained from CRSP, and institutional ownership data are obtained from the Thomson Reuters 13F database. I exclude observations that are lacking necessary data from Compustat, CRSP, I/B/E/S, or Thomson Reuters, yielding a final sample of 1,067 firms and 14,222 firm-quarter observations.

Appendix Table A2 presents the frequency distribution of tweets and firm-quarter observations by calendar quarter. In my sample, the frequency of tweets over time is relatively flat, while the number of firm-quarter observations increases over the sample. This pattern is to be expected, because some Twitter users in the sample were not active at the start of the sample period.

There is considerable heterogeneity across firms' Twitter accounts, which suggests that the effect of tweeting may vary by firm. To address this concern I use firm fixed effects and standard errors clustered by firm. I also control for the number of retweets when measuring the impact of firm tweets. Appendix Table A3 presents descriptive statistics related to tweet characteristics.

In order to study the high-frequency dynamics of stock returns around earnings disclosures on Twitter, I use minute-level price data from Bloomberg. Due to data availability this dataset

[5] A reply tweet is a public tweet directed at a specific person. Reply tweets do not appear in the feeds of everyone following the firm; rather, they appear only in the feed of the specific user to whom the firm is replying and in the feed of anyone else who follows both the replying firm and the user receiving the reply. A retweet is the reposting of another Twitter user's tweet on the firm's own profile. Unlike reply tweets, retweets appear in the feed of everyone who is following the firm that reposts the tweet. However, the retweet itself is not original content created by the firm.

spans from November 2019 through July 2021. Tweets are matched to the price data using the same procedure as outlined above.

To further investigate fundamental information acquisition, I utilize the SEC's EDGAR log file dataset. This dataset is a collection of web server log files that allows researchers to study firm-specific web traffic of individuals downloading SEC filings. EDGAR is the central repository for all mandatory SEC filings, and the daily-level EDGAR search volume for each firm is a direct measure of investors' fundamental information acquisition. EDGAR log file data are obtained from James Ryans's webpage.[6]

D. Identifying financial news tweets

One of the primary challenges underlying the research design is the detection of financial news tweets. Following prior research, I use textual analysis to identify these tweets (see, for example, Bartov et al. (2017) and Jung et al. (2018)). I use a classification scheme based on a dictionary of key words and phrases; each tweet is considered earnings news if it contains two or more of the terms found in this dictionary.[7]

Using this textual classification approach, I identify 19,148 tweets (5,549 firm-quarters, 783 unique firms) that contain information directly related to earnings announcements. Examples of financial news tweets in the sample are provided in Appendix Figure 5. As one would expect, financial news tweets are concentrated around earnings announcement periods. The number of financial news tweets in a 10-day window around the announcement represents, on average, approximately one-fourth of all tweets in that period.[8]

Figure 1 depicts the total number of financial news tweets that are posted each hour in the 48 hours before and after earnings are announced. On average, financial news tweets reach their peak numbers the two hours after a quarterly earnings announcement; however, a considerable portion of

[6]The summarized EDGAR log files used in this paper are available for academic use at http://www.jamesryans.com.

[7]The dictionary of key words and phrases can be found in the Appendix.

[8]Summary statistics are provided in Appendix Table A4.

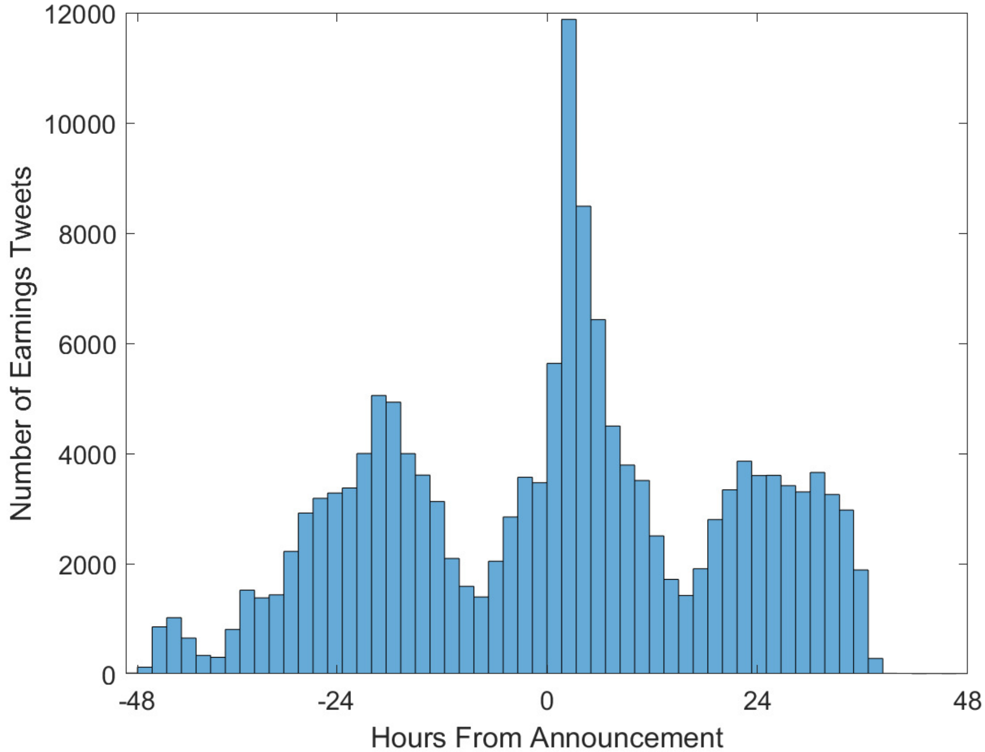


Figure 1: Tweeting around earnings announcements. This figure depicts the relationship between the number of financial news tweets and the number of hours away from firms' earnings announcements. All tweets included in the figure were posted during the sample period (Jan. 2014 through Dec. 2017) by S&P 1500 firms and meet basic minimum word requirements to be considered financial news tweets.

the distribution of financial news tweets are posted in the days before and after the announcement.

E. Measuring network impact

Because Twitter is an interactive network, it is essential to consider the diffusion of tweets in the network when measuring the relative impact of individual tweets. When a firm posts a tweet, this message is immediately accessible to the firm's followers. These followers have the option to interact with the tweets; if the tweet is retweeted or liked by one of the firm's followers, then the tweet can be seen by both the firm's followers and the other user's followers. As the process of retweeting and liking continues, a tweet can potentially spread through the entire network.

To capture these network effects, I measure the impact of firms' financial news tweets (*FinNew-*

sTweetImpact) in two ways. First, I use the IHS transformation of the number of financial news tweets as a naive proxy for the impact of tweets in the network. Second, I multiply the IHS transformation of the number of financial news tweets by the IHS transformation of the number of financial news retweets to further capture the diffusion of tweets in the network..

3. Theoretical Setting

An extensive theoretical literature has studied when and why limited voluntary disclosure is likely to occur. The "unraveling result" established by Grossman and Hart (1980), Grossman (1981), and Milgrom (1981) identified conditions under which firms voluntarily disclose all private information in equilibrium. One of the most fragile conditions is that investors must be fully informed about the manager's information endowment. If investors are uncertain whether managers have private information, the managers may withhold information in equilibrium Dye (1985).

Disclosures made through social media channels must be truthful and accurate, as do all other forms of disclosure by regulated firms. However it is left up to the managers' discretion whether or not to disclose information on social media at all. Therefore, a firm's choice to disclose information on social media may reflect the strategic decisions of managers who have a material interest in the reaction of the market to new information. The strategic disclosure model of Shin (2003) formalizes the concept that "although the manager has to tell the truth, he cannot be forced to tell the whole truth" (p. 108). Goto et al. (2009) extend Shin's analysis to investigate the effects of strategic disclosure on the time-series behavior of stock returns in comparison to the effects of full disclosure.

A. Model

I examine the effects of voluntary strategic disclosure after earnings announcement events using the model introduced by Goto et al. (2009). In my setting the success of each firm depends on N independent and identical dimensions, where exante each dimension of the business succeeds with probability r and fails with probability $1 - r$. There are three dates, 0, 1, and 2. At date 0, a firm announces its quarterly earnings results to the market and investors update their prior on r. If a firm announces earnings below (above) market expectations, the probability that a business dimension succeeds becomes $r_l < r$ ($r_h > r$). Each dimension of the firm's business is realized by date 1 with probability θ and observed by the firm's manager. At date 1 managers observe s number of successes and f number failures, and have the opportunity to voluntarily disseminate s' successes and f' failures (where $0 \leq s' < s$ and $0 \leq f' < f$), presumably via social media. It is important to note that the earnings announcement at $t = 0$ is a mandatory disclosure, while the use of social media at $t = 1$ is not required by regulators. By the final date, the outcomes of all business dimensions become common knowledge and the firm is liquidated. The liquidation value of the firm depends on the total number of successes (k) and failures ($N - k$). Each successful business dimension corresponds to a jump up in a binomial pricing tree that increases the liquidation value by a factor of u, and each failed project corresponds to a jump down by a factor of d.

At date 1 there is asymmetric information between managers and investors. If the successes or failures of a business dimension is realized before date 1, the manager observes the outcome, however, investors only observe the disclosure by managers. It is important to point out that managers cannot lie about the success of a project (hence $0 \leq s' < s$), but they are free to disclose successes and withhold disclosure of failures if they deem it favorable. The idea is that a manager's disclosures can be verified at a later date, and therefore an outside party, such as a court, can impose a penalty if a past disclosure is found to be untrue. That said, the amount of private information the manager has at date 1 is not verifiable, and therefore a manager is free to withhold the outcomes of projects if those outcomes are unfavorable. Investors know that the manager is informed with some probability, therefore, if the manager chooses not to disclose information it could either be that they are uninformed or that the information is bad.

I examine two manager strategies: one in which the manager follows a strategic disclosure policy (only disclosing success at date 1), and another in which the manager follows a full disclosure policy (disclosing successes and failures at date 1). I do not consider the strategy of non-disclosure since there is always an incentive to deviate at $t = 1$ by disclosing some successes, and hence this strategy is never supported in equilibrium.

Under reasonable parametric assumptions the model shows that the jump in expected returns associated with using a strategic disclosure strategy is decreasing in r. Since investors update their prior on r at date 0, this means that the expected return when using a strategic disclosure strategy is higher for firms which announce a negative earnings surprise at date 0 than for firms that announce positive earnings surprise (see Appendix section E for details). The intuition is that the marginal benefit of strategically disseminating information on social media is higher for firms that are less likely to have good news to disclose. Hence, the model predicts that stock prices will rise more for firms that follow a strategic disclosure policy following a relatively poor earnings announcement.

The probability that a manager is informed about the outcome of a business dimension at the interim date captures the relative information asymmetry. When θ increases, managers are more likely to observe the outcomes of their business dimensions at $t = 1$ and asymmetric information between managers and investors is higher. Therefore, θ can be thought of as a measure of relative information asymmetry. Under reasonable parametric assumptions the model predicts that the jump in expected returns associated with using a strategic disclosure strategy is increasing in θ. The idea is that the marginal benefit of strategically disseminating information on social media is higher for firms with high levels of information asymmetry.

4. Empirical Design and Results

A. Disclosure strategy

I begin my empirical analysis by investigating the relevant drivers of tweeting about financial news. To test whether firms have a full disclosure policy or whether the disclosure depends on the extent that a firm is revealing positive or negative news, I estimate the following regression:

$$\begin{aligned} FinNewsTweets_{i,t} = \alpha + \beta_1 UnexpectedEarnings_{i,t} \\ + \beta_2 X_{i,t} + \theta_i + \psi_t + \varepsilon_{i,t}. \end{aligned} \tag{1}$$

The control variables, $X_{i,t}$, include $StockMarketIndex_{i_t}$, $Size_{i,t}$, $B/M_{i,t}$, $Analysts_{i,t}$, $Q4_{i,t}$, $Loss_{i,t}$, $InstitutionalOwnership_{i,t}$, $TwitterNetworkSize_i$, and $VerifiedTwitterAccount_i$. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. Appendix Table A4 reports summary statistics on the variables used to estimate equation (1). All variables are defined in detail in Appendix Table A5.

Table 1 displays the coefficient estimates. In columns (1) and (2) I report the results of estimating equation (1) with a probit model. The dependent variable, $FinNewsTweetDummy_{i,t}$, is a binary outcome variable equal to 1 if a firm tweets about earnings over the three-day window [–1, +1] around the earnings announcement and zero otherwise. The variable $SUE_{i,t}$ (standardized unexpected earnings) is the firm's actual earnings minus the analyst consensus forecast of earnings, standardized by the standard deviation of analyst forecasts. This variable captures the "surprise" aspect of the earnings news.

In columns (1) and (2), SUE and $NegativeSurprise$ have significant coefficients at 1%. This indicates that the choice to tweet about financial news does in fact depend on the extent to which a firm is revealing positive or negative news. This result is in line with the finding of Jung et al. (2018) and suggests that on average firms tend to strategically disclose news on social media.

The probit specification in Table 1 enables me to compare characteristics across firms that impact the likelihood of tweeting about financial news. I note several interesting patterns. Firms that belong to the S&P 500 index (large-cap)–that is, large, well-known firms—are more likely to tweet about financial news. It is important to note that this result holds despite controlling for the size of firms. Firms with lower book-to-market values are also more likely to tweet about financial news. On average, technology companies and other companies in industries that do not have a lot of physical assets tend to have low book-to-market ratios.

One concern about using a probit model is that fixed effects cannot be controlled for. Therefore I re-estimate equation (1) using an OLS model with firm and quarter fixed effects. In columns (3) and (4) I report the results of estimating equation (1) with an OLS model. The dependent variable, $FinNewsTweetCount_{i,t}$, is the number of financial news tweets over the three-day window [–1, +1] around the earnings announcement. All other variables remain the same.

In columns (3) and (4) of Table 1 both SUE and $NegativeSurprise$ have significant coefficients. These results are consistent with those of the probit specification and reaffirm that the choice to tweet about financial news depends on the extent to which a firm is revealing positive or negative news. *Loss* changes sign when I control for firm and quarter fixed effects–this indicates that a firm is less likely to tweet about financial news when its net income is negative than when its net income is positive.

B. Tweeting and announcement returns

My primary research question focuses on whether there is a link between corporate information dissemination on social media and stock returns. To answer this question I investigate the relationship between announcement returns and financial news tweets. If managers follow a strategic disclosure scheme in equilibrium, then on average one would expect to see negative earnings surprise to be met with a larger increase in returns if managers strategically tweet about financial news. To test this hypothesis I estimate the following model:

$$\begin{aligned} CAR_{i,t} = \alpha &+ \beta_1 NegativeSurprise_{i,t} + \beta_2 FinNewsTweetImpact_{i,t} \\ &+ \beta_3 NegativeSurprise_{i,t} \times FinNewsTweetImpact_{i,t} \\ &+ \beta_4 X_{i,t} + \theta_i + \psi_t + \varepsilon_{i,t}. \end{aligned} \tag{2}$$

In equation (2), the dependent variable, $CAR_{i,t}$, is the Carhart (1997) cumulative abnormal return for firm i over the three-day window [–1, +1] around the quarterly earnings announcement.

Table 1: When Firms Tweet about Financial News

In this table I test whether firms have a full disclosure policy or whether their disclosure depends on the extent to which they are revealing positive or negative news. In columns (1) and (2) I estimate equation (1) with a probit model, and the dependent variable is $FinNewsTweetDummy_{i,t}$, a binary outcome variable equal to 1 if a firm tweets about earnings over the three-day window [-1, +1] around the earnings announcement and zero otherwise. In columns (3) and (4) I estimate equation (1) using an OLS model with firm and quarter fixed effects, and the dependent variable is $FinNewsTweetCount_{i,t}$, the number of financial news tweets over the three-day window [–1, +1] around the earnings announcement. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. See Appendix Table A5 for variable definitions. Standard errors are clustered at the firm level and reported in parentheses beneath the coefficient estimates. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Probit		**Fixed Effects**	
	Fin News Tweet Dummy		*Fin News Tweet Count*	
	(1)	(2)	(3)	(4)
Std. Unexpected Earnings	0.017***		0.007***	
	(0.005)		(0.003)	
Negative Surprise		-0.096***		-0.037*
		(0.034)		(0.021)
Additional Tweet Count	0.004***	0.004***	0.008***	0.008***
	(0.002)	(0.002)	(0.003)	(0.003)
SP500	0.330***	0.330***	-0.059	-0.055
	(0.107)	(0.107)	(0.237)	(0.238)
SP600	-0.207**	-0.209**	-0.012	-0.013
	(0.085)	(0.085)	(0.073)	(0.074)
Size	0.178***	0.177***	-0.012	-0.010
	(0.031)	(0.031)	(0.099)	(0.099)
BM	-0.288***	-0.288***	-0.156	-0.156
	(0.110)	(0.110)	(0.266)	(0.266)
Loss	0.216***	0.210***	-0.124**	-0.127***
	(0.069)	(0.069)	(0.048)	(0.049)
Q4	-0.030	-0.032	-0.029	-0.031
	(0.025)	(0.025)	(0.025)	(0.025)
Analysts	0.005	0.004	-0.005	-0.005
	(0.005)	(0.005)	(0.003)	(0.003)
Institutional Ownership	-0.098	-0.091	0.173	0.170
	(0.167)	(0.168)	(0.288)	(0.288)
Twitter Network Size	-0.209***	-0.210***		
	(0.025)	(0.025)		
Verified Twitter Account	-0.017	-0.014		
	(0.099)	(0.099)		
Firm FE	No	No	Yes	Yes
Quarter-Year FE	No	No	Yes	Yes
Adjusted (Pseudo) R^2	(0.098)	(0.097)	0.624	0.624
Observations	14,045	14,045	14,222	14,222

$NegativeSurprise_{i,t}$ is an indicator variable equal to one if firm i misses its analyst consensus forecast in quarter t, and zero otherwise. $FinNewsTweetImpact_{i,t}$ captures the extent to which a firm i tweets about their quarterly earnings announcement over the three-day window [–1, +1] around the announcement. First, I use the inverse hyperbolic sine (IHS) transformation of the number of financial news tweets. Second, I multiply the IHS transformation of the number of financial news tweets by the IHS transformation of the number of financial news retweets to capture diffusion of information in the network. $NegativeSurprise_{i,t} \times FinNewsTweetImpact_{i,t}$ is an interaction term; this variable helps capture the impact of a firm's financial news tweets, given that the firm misses its consensus forecast.

Beating analysts' forecasts of earnings is a concept well studied by researchers. The literature has shown that the market response to earnings surprises is asymmetric. Skinner and Sloan (2002) find that the price reaction to a negative surprise tends to be larger in magnitude than the price reaction to a positive surprise. Moreover, there is a large jump in density when going from firms with a negative surprise of 1 cent to those having no surprise at all, which highlights the high cost of missing analysts' expectations (Matsumoto, 2002).

Furthermore, the premium from having no surprise or a positive surprise even exists in the cases in which the forecasted earnings target is likely to have been achieved through earnings or expectations management Bartov et al. (2002). Given the asymmetry in the market response to positive and negative earnings announcements, I choose to interact $FinNewsTweetImpact_{i,t}$ with the dummy variable $NegativeSurprise_{i,t}$, rather than with the continuous variable $SUE_{i,t}$.

The control variables, $X_{i,t}$, include $Size_{i,t}$, $B/M_{i,t}$, $Analysts_{i,t}$, $SUE_{i,t}$, $Q4_{i,t}$, $Loss_{i,t}$, and $AdditionalTweetImpact_{i,t}$. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. Appendix Table A4 reports summary statistics on the variables used to estimate equation (2). All variables are defined in detail in Appendix Table A5.

In Table 2, equation (2) is estimated using firm and quarter-year fixed effects. For firms that miss their analyst consensus forecast, the effect of tweeting about earnings in column (2) is 0.866 (0.100 + 0.766), which the F-test shows is significant at 1%. The coefficient estimate for

FinNewsTweetImpact in column (2) is 0.100 and is statistically insignificant, meaning that for firms that meet or beat their analyst consensus forecast, tweeting about earnings is not associated with a change in the announcement return.

The within-group estimates suggest that when the same firm tweets about earnings over different quarters, tweeting has an asymmetric effect on announcement returns depending on whether the firm has a positive or negative earnings surprise. Firms with negative surprises have higher announcement returns when they tweet about earnings news. These results are robust to the measurement of *FinNewsTweetImpact* in columns (3) and (4).

One implication of strategic disclosure is that the jump in expected returns associated with using a strategic disclosure strategy is stronger for firms which announce a negative earnings surprise at date 0. In line with this prediction, I find firms that tweet about financial news following a negative earnings surprise have higher abnormal returns. These results establish a link between corporate information dissemination on social media and stock returns and support the theoretical predictions outlined in section 3.

Firm level ownership and visibility

As discussed in Section 3., incentives for strategic dissemination may be related to a firm's level of information asymmetry between managers and investors. In this subsection, I test whether variation in firm characteristics associated with information asymmetry can help explain higher announcement returns when firms tweet about a negative earnings surprise event.

Information asymmetry often corresponds with the level of investor sophistication. Unsophisticated investors tend to have higher costs of information acquisition and processing and therefore are relatively less informed than sophisticated investors. As discussed in Section 3., my theoretical framework predicts that the jump in expected returns, when going from a full disclosure policy to a strategic one, is increasing in the relative level of information uncertainty. To proxy for the level of sophistication in a firm's investor base I use the percentage of shares outstanding owned by retail

Table 2: Tweeting and Announcement Returns

This table shows the relationship between cumulative abnormal returns and firms' tweeting behaviors. In columns (1) and (2) *FinNewsTweetImpact* is measured as as *FinNewsTweets* and in columns (3) and (4) as *FinNewsTweets*Retweets*. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. Standard errors are clustered at the firm level and reported in parentheses beneath the coefficient estimates. *, **, and *** indicates significance at the 10%, 5%, and 1% levels, respectively.

	$CAR_{-1,+1}$			
	(1)	(2)	(3)	(4)
Negative Surprise	-3.447***	-3.762***	-3.447***	-3.648***
	(0.189)	(0.215)	(0.189)	(0.201)
Fin News Tweet Impact	0.280*	0.100	0.071	-0.010
	(0.145)	(0.153)	(0.073)	(0.075)
Negative Surprise × *Fin News Tweet Impact*		0.766***		0.468***
		(0.229)		(0.106)
SUE	0.465***	0.465***	0.465***	0.464***
	(0.028)	(0.028)	(0.028)	(0.028)
Additional Tweet Impact	0.040	0.041	0.047	0.051
	(0.081)	(0.081)	(0.081)	(0.034)
Residual ESV (Ryans)	-0.444	-0.443	-0.444	-0.450
	(0.287)	(0.288)	(0.287)	(0.288)
Size	-2.473***	-2.469***	-2.477***	-2.477***
	(0.409)	(0.410)	(0.410)	(0.411)
Loss	-1.661***	-1.644***	-1.672***	-1.644***
	(0.304)	(0.304)	(0.304)	(0.304)
BM	7.065***	7.067***	7.045***	7.056***
	(0.750)	(0.748)	(0.747)	(0.746)
Q4	0.311	0.299	0.308	0.296
	(0.203)	(0.203)	(0.204)	(0.203)
Firm FE	Yes	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes	Yes
F-test statistic: $\beta 2 + \beta 3 = 0$		14.054		15.292
F-test p-value		0.000		0.000
No. of firms	1067	1067	1067	1067
Adjusted R^2	0.155	0.156	0.155	0.156
Observations	14,222	14,222	14,222	14,222

investors. Using institutional ownership holdings from Thomson Reuters, I compute the percentage of retail investors as 100 percent less the percentage of shares outstanding owned by institutions. I sort firms into low (high) investor sophistication categories if their retail ownership as a percent of shares outstanding is above (below) the sample median of 16 percent. I re-estimate equation (2), with the addition of a triple interaction term *High Retail* $\times$ *Neg Surprise* $\times$ *Fin News Tweets*. The results are provided in column (1) of Table 3. The positive relationship between tweeting after a negative earnings announcement and daily returns is stronger in firms with high retail ownership.

Information asymmetry may also correspond with how visible a firm is to investors. High visibility firms are more likely to receive broad coverage through traditional channels like business press and analysts reports. Therefore the more visible a firm is, the easier it is for investors to access information about that firm. To proxy for visibility I use two measures, analysts coverage and size squared. Equity analysts are important information intermediaries that provide investors with detailed financial analyses and recommendations on whether to buy, hold, or sell a particular investment. I use the number of analysts following each firm as a measure of visibility and information available to investors. Very large firms also tend to be the visible and well know. By using size squared I am able to capture non-literariness in the model for the largest and most well known firms. I re-estimate equation (2), with the addition of the triple interaction terms *Analysts*$\times$ *Neg Surprise* $\times$ *Fin News Tweets* and $Size^2\times$ *Neg Surprise* $\times$ *Fin News Tweets*, the results are receptively provided in columns (2) and (3) of Table 3. The positive relationship between tweeting after a negative earnings announcement and daily returns is stronger in highly visible firms.

Overall, the results in Table 3 suggest that, as predicted, the dissemination of public information on social media matters more for retail investors, which tend to have higher costs of information acquisition and processing. Further social media is especially important for large and highly visible firms, which suggests that having Twitter account is not a substitute for overall visibility but rather a complement to it.

Table 3: Retail Investor Ownership and Visibility

This table shows the relationship between cumulative abnormal returns and firms' tweeting behaviors. In column (1) *Neg Surprise* × *Fin News Tweets* is interacted with the dummy variable *High Retail*, which is equal to one if a firm's average institutional ownership is below the sample median. In column (2) *Neg Surprise* × *Fin News Tweets* is interacted with the variable *Analysts*, which measures the number of analysts following each firm in a given quarter. In column (3) *Neg Surprise* × *Fin News Tweets* is interacted with the variable $Size^2$, which is the quadratic term of *Size* and is a proxy for vary large firms. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. Standard errors are clustered at the firm level and provided in parentheses beneath the coefficient estimates. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

	$CAR_{-1,+1}$		
	(1)	(2)	(3)
Negative Surprise	-5.592***	-5.591***	-5.574***
	(0.165)	(0.165)	(0.165)
Fin News Tweets	0.164	0.177	0.210
	(0.158)	(0.158)	(0.158)
Neg Surprise ×*Fin News Tweets*	0.293	0.063	-1.306***
	(0.297)	(0.328)	(0.498)
High Retail × *Neg Surprise* × *Fin News Tweets*	0.765***		
	(0.283)		
Analysts × *Neg Surprise* × *Fin News Tweets*		0.062***	
		(0.023)	
$Size^2$× *Neg Surprise* × *Fin News Tweets*			0.023***
			(0.005)
Additional Tweets	0.027	0.023	0.023
	(0.084)	(0.084)	(0.084)
Residual ESV	-0.450	-0.455	-0.459
	(0.311)	(0.311)	(0.311)
Size	-2.401***	-2.401***	-2.436***
	(0.352)	(0.352)	(0.351)
Loss	-1.902***	-1.882***	-1.890***
	(0.234)	(0.234)	(0.234)
BM	7.140***	7.125***	7.136***
	(0.450)	(0.450)	(0.450)
Q4	0.203	0.224	0.194
	(0.204)	(0.204)	(0.204)
Firm FE	Yes	Yes	Yes
Year-quarter FE	Yes	Yes	Yes
No. of firms	1067	1067	1067
Adjusted R^2	0.127	0.127	0.128
Observations	14,222	14,222	14,222

C. *Tweeting and announcement returns: A high-frequency analysis*

Twitter was made to share news, content, and information in real time. This platform enables firms to share financial news with their social network, and individuals to have instantaneous access to that information. The speed of information flow on Twitter creates a unique setting in which to study the possible reactions of investors to tweets about financial news. In this section I study the high-frequency dynamics of stock returns around financial news disclosures on Twitter.

My novel dataset of financial news tweets enables me to measure the exact time of information disclosures on Twitter. For each tweet in the dataset, I compile the minute-level return data in the 60-minute window around the tweet and estimate the following model at the minute level:

$$Y_t = (\alpha_{pre} + \alpha_{t>t^*}) + (\beta_{pre} + \beta_{t>t^*})t + (\gamma_{pre} + \gamma_{t>t*})t^2 + \varepsilon_t. \tag{3}$$

In equation (3), t^* is the time of a financial news tweet. The dependent variable Y_t is either the average abnormal cumulative return obtained from buying equities 30 minutes before a tweet and holding them for 60 minutes, or the average minute-level trading volume. The model is a quadratic function of time that includes dummy variables to account for post-announcement jumps in intercept and slope. I test the null assumption that there is no difference in the post-announcement window, $H_0 : \alpha_{t>t^*} = \beta_{t>t^*} = \gamma_{t>t*}$. If the null hypothesis is rejected, two separate functions are fit, one before t^* and one after. The results are aggregate across firms and quarters. Quarters are divided into two groups, positive and negative earnings surprise quarters, according to the standard definition based on analyst consensus forecasts.

Equity Returns

Figure 2 depicts the results visually. In the upper panel of Figure 2 I consider all tweets about financial news within a three-day window around earnings announcement events, in quarters in which earnings are reported above the analyst consensus forecast. The figure depicts the average cumulative return in excess of the same average measured in a control sample, defined using the same time of day in a matched quarter in which tweets do not occur. The matching procedure controls for the level of surprise, relative to the analyst consensus forecast, and the time of day. This strategy allows me to control for the common trend, which is generally upward sloping on days with a positive surprise and downward sloping on days with a negative surprise.

For positive-surprise events (panel (a)), abnormal cumulative returns tend to increase before the tweet and then decline slightly upon the announcement. Turning our attention to negative-surprise events (panel (b)), abnormal cumulative returns tend to appreciate before the tweet and then further appreciate upon the announcement. This observation suggests that the release of financial news on Twitter may help equities during periods surrounding poor earnings announcements.

These results confirm the previous analysis using minute-level data rather than daily data. In Table 4 you can the coefficient estimates based on equation (3).

Trading Volume

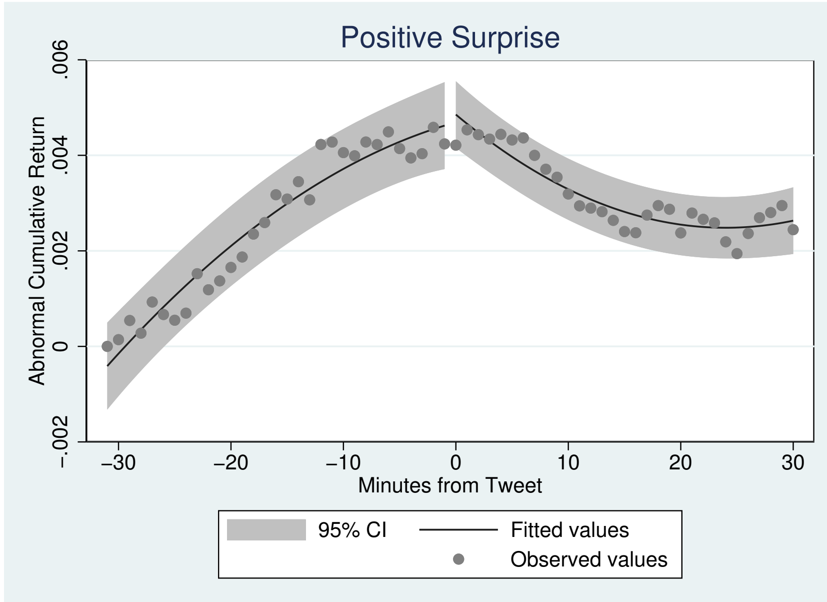
Figure 3 shows the high-frequency results for trading volumes. In panel (a) I consider all tweets about financial news within a three-day window around earnings announcement events, in quarters in which earnings are reported *above* the analyst consensus forecast. Instead, in panel (b) I consider quarters in which earnings are reported *below* the analyst consensus forecast.

For positive-surprise events (panel (a)), trading volumes tend to slightly decrease before the tweet and then increase for about 20 minutes immediately after the announcement. For negative-surprise events (panel (b)), trading volumes tend to slightly increase before the tweet and then continue to increase for about 15 minutes immediately after the announcement. This observation suggests that investors do in fact trade on the information released via Twitter. In both panels the concave shape of the function in the 30 minutes following the tweets suggests that trading responds quickly to tweets about financial news but the tweets have only a transitory effect.

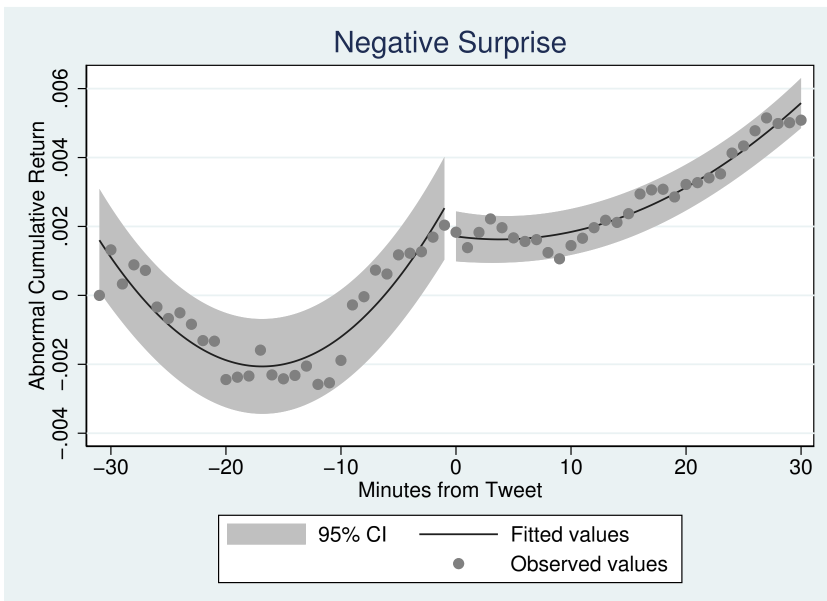
Table 4: High Frequency Returns: Regression Analysis

This table shows the coefficient estimates based on the estimation of equation (3). The sample is split into positive and negative news surprise disclosures. In columns (1) and (2) the dependent variable is the cumulative abnormal return obtained from buying equities 30 minutes before a tweet and holding them for 60 minutes. In columns (3) and (4) the dependent variable is the minute-level trading volume from 30 minutes before a tweet to 30 minutes after. Standard errors are reported in parentheses beneath the coefficient estimates. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Abnormal Cumulative Return		Trading Volume	
	Pos. Surp.	Neg. Surp.	Pos. Surp.	Neg. Surp.
Post	1.6070***	0.6947***	-51243.4980***	-26736.1007**
	(0.196)	(0.180)	(16205.046)	(10918.360)
t	0.0270***	-0.0554***	-447.2025	194.7015*
	(0.003)	(0.007)	(457.736)	(102.216)
Post $\times$ t	-0.0731***	0.0138	2613.6193***	1141.6288**
	(0.009)	(0.010)	(832.772)	(493.152)
t^2	-0.0003***	0.0018***	10.2321	-2.6874
	(0.000)	(0.000)	(11.715)	(3.237)
Post $\times$ t^2	0.0007***	-0.0012***	-31.9605**	-10.7904*
	(0.000)	(0.000)	(13.808)	(6.066)
Constant	-0.0679***	0.2137***	21071.2402***	9364.6844***
	(0.022)	(0.055)	(3891.076)	(639.326)
R^2	0.930	0.950	0.501	0.511
Observations	62	62	62	62

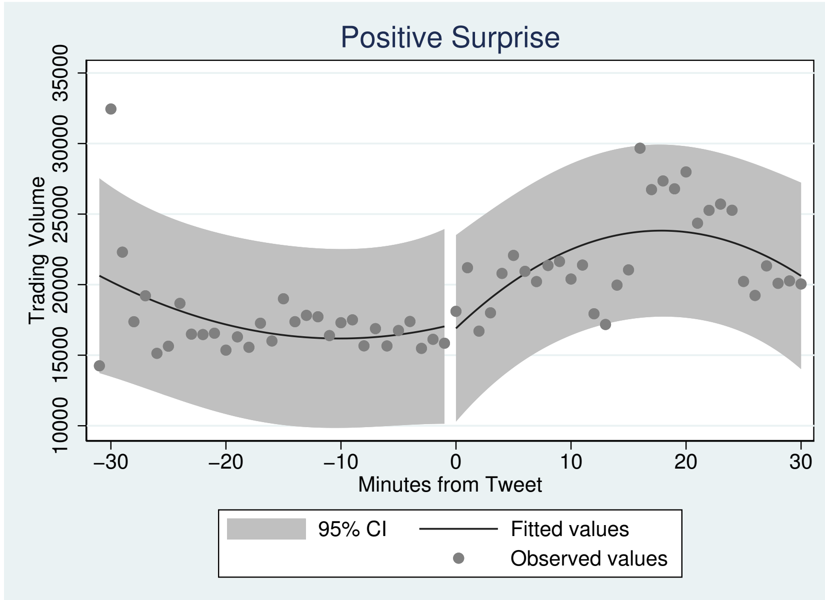


(a)

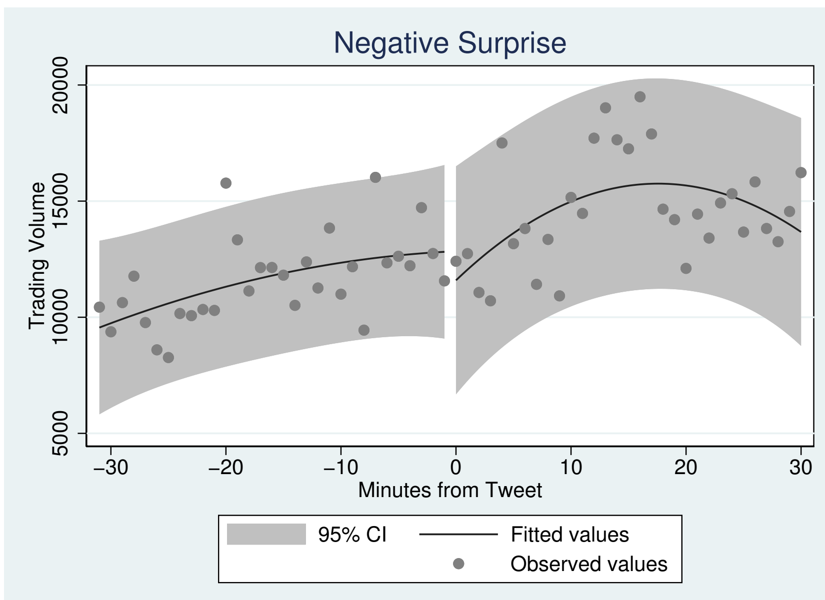


(b)

Figure 2: High frequency equity returns: This figure depicts the average abnormal cumulative return obtained from buying equities 30 minutes before a tweet and holding them for 60 minutes. The abnormal cumulative return is defined as the average cumulative return in excess of the control sample, where the control sample is defined using a matched firm-time where tweets do not occur. The matching procedure controls for the level of surprise, relative the analyst consensus forecast, and the time of day. The sample is split into positive news and negative news disclosures, panel a (panel b) depicts returns around positive (negative) news. The solid lines and shaded areas are based on the estimation of equation (3) where $t^* = 0$ is the time of an earnings related tweet. The null assumption that there is no difference in the post-announcement window, $H_0 : \alpha_{t>t^*} = \beta_{t>t^*} = \gamma_{t>t*}$, and if the null hypothesis is not rejected a continuous quadratic function if fit. Standard errors are estimated at 95%. Returns are in raw log units.



(a)



(b)

Figure 3: High frequency trading volumes: This figure depicts the average trading volume from 30 minutes before a tweet to 30 minutes after a tweet. The sample is split into positive news and negative news disclosures, panel a (panel b) depicts the trading volume around positive (negative) news. The solid lines and shaded areas are based on the estimation of equation (3) where $t^* = 0$ is the time of an earnings related tweet. The null assumption that there is no difference in the post-announcement window, $H_0 : \alpha_{t>t^*} = \beta_{t>t^*} = \gamma_{t>t*}$, and if the null hypothesis is not rejected a continuous quadratic function if fit. Standard errors are estimated at 95%.

D. Tweeting and fundamental information acquisition

Does tweeting encourage fundamental information acquisition? To test this question I use a novel dataset that tracks all web traffic on the SEC's EDGAR servers. The SEC has assembled a log file which records each user request to acquire a specific filing from EDGAR. This dataset allows me to analyze investor acquisition of specific financial disclosures and study the relationship between information acquisition and a firm's tweeting behavior by estimating the following regression:

$$\begin{aligned} ESV_{i,t} = \alpha &+ \beta_1 FinNewsTweetImpact_{i,t} + \beta_2 NegaticeSurprise_{i,t} \\ &+ \beta_3 X_{i,t} + \theta_i + \psi_t + \varepsilon_{i,t}. \end{aligned} \tag{4}$$

In equation (4), the dependent variable, $ESV_{i,t}$, is the daily EDGAR Search Volume from the SEC's web server log file data for firm i over the three-day window [–1, +1] around the quarterly earnings announcement. Since the log files must be filtered to remove downloads by computer programs, I use two methods for counting human views in the EDGAR log files developed by Ryans (2017) and Loughran and McDonald (2017). $FinNewsTweetImpact_{i,t}$ captures the extent to which firm i tweets about its earnings announcement over the three-day window [–1, +1] around the earnings announcement.

The control variables, $X_{i,t}$, include $Size_{i,t}$, $B/M_{i,t}$, $Analysts_{i,t}$, $SUEi,t$, $Q4_{i,t}$, $Loss_{i,t}$, and *AdditionalTweetImpact*$_{i,t}$. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. Appendix Table A4 reports summary statistics on the variables used to estimate equation (2). All variables are defined in detail in Appendix Table A5.

In Table 5, equation (4) is estimated using firm and quarter-year fixed effects. The coefficient estimates for *FinNewsTweetImpact* are statistically significant at the 5% level or higher in all columns except (3), indicating that in most specifications tweeting about financial news is associated with more fundamental information acquisition by individual investors. In columns (2) and (4) I include the interaction term $NegativeSurprise \times FinNewsTweetImpact_{i,t}$; this specification reveals that only firms' tweeting about a positive earnings surprises is associated with higher EDGAR search

Table 5: Tweeting and Fundamental Information Acquisition

This table shows the relationship between the dependent variable, EDGAR search volume, and firms' tweeting behavior. The dependent variable, ESV, is the daily EDGAR Search Volume from the SEC's web server log file over the three-day window [-1, +1] around the quarterly earnings announcement. In columns (1) and (2) I follow log file cleaning procedure developed by Loughran and McDonald (2017), and in columns (3) and (4) I follow Ryans (2017). $FinNewsTweetImpact$ is the IHS transformation financial news tweets in the three-day window [-1, +1] around the earnings announcement. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. See Appendix table A5 for variable definitions. The regression is estimated using OLS with robust standard errors. Standard errors are provided in parentheses beneath the coefficient estimates. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

	Edgar Search Volume			
	(1)	(2)	(3)	(4)
Fin News Tweet Impact	0.025**	0.030***	0.015	0.022**
	(0.011)	(0.011)	(0.010)	(0.011)
Negative Surprise	0.035***	0.044***	0.028***	0.039***
	(0.010)	(0.011)	(0.008)	(0.010)
*Neg Surp*Fin News Tweet Impact*		-0.021*		-0.027**
		(0.012)		(0.011)
Additional Tweet Impact	-0.008	-0.008	-0.013**	-0.013**
	(0.006)	(0.006)	(0.006)	(0.006)
Size	0.261***	0.260***	0.271***	0.270***
	(0.035)	(0.035)	(0.031)	(0.031)
Loss	0.052***	0.051***	0.037**	0.036**
	(0.017)	(0.017)	(0.015)	(0.015)
BM	-0.009	-0.009	-0.021	-0.022
	(0.038)	(0.039)	(0.037)	(0.037)
Q4	0.044***	0.045***	0.020	0.020
	(0.015)	(0.015)	(0.013)	(0.013)
Analysts	0.003***	0.003***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.001)
Firm FE	Yes	Yes	Yes	Yes
Quarter-year FE	Yes	Yes	Yes	Yes
F-test statistic: $\beta2 + \beta3 = 0$		0.506		0.192
F-test p-value		0.477		0.661
No. of firms	1,030	1,030	1,030	1,030
Adjusted R^2	0.810	0.810	0.809	0.809
Observations	12,484	12,484	12,484	12,484

volumes. Instead, firms' tweeting about negative earnings surprises is not associated with a change in EDGAR search volumes.

E. *Tweeting and the speed of information diffusion*

Can firm-initiated tweets increase the speed of information diffusion? Momentum in returns has been explained theoretically and empirically by gradual diffusion of information (Hong and Stein (1999), Hong et al. (2000)). Momentum in stock returns is a longstanding empirical fact; that is, securities which have performed well over the prior 6-12 months continue to outperform relative to those that did poorly, for the next 3-12 months Jegadeesh and Titman (1993).

If tweeting about earnings news increases the speed of information diffusion to the market, then momentum in returns should decrease. To test this prediction I estimate the following regression:

$$Momentum_i = \alpha + \beta_1 EarningsTweetQuarters_i + \beta_2 X_i + \varepsilon_i. \tag{3}$$

In equation (3), the dependent variable, $Momentum_i$, is a proxy for momentum as it is defined in the empirical asset pricing literature (cf. Jegadeesh and Titman (1993)). $Momentum_i$ is measured as the correlation between the series $ExRet_{i,t}$ and the lagged series $ExRet_{i[t-12,t-2]}$, where $ExRet_{i,t}$ is the monthly excess return of firm i. $EarningsTweetQuarters_i$ is the proportion of quarters in which a firm tweets about earnings news over the sample period, January 2014 through December 2017. I construct $Momentum_i$ using $t \in \{\text{January 2014,..., December 2017}\}$ to match the sample period.

The controls, X_i, include $Size_i$, B/M_i, and $Analysts_i$ and are measured using the average value over the sample period. Appendix Table A4 presents the descriptive statistics for the variables used to estimate equation (3). All variables are defined in detail in Appendix Table A5.

In Table 6 *Momentum* is calculated using excess returns relative to 90-day T-bills (columns (1) and (2)) and using Fama-French three-factor excess returns (columns (2) and (4)). In columns (1) and (2) the coefficients are estimated using the full sample of firms, and the coefficient estimates for *FinNewsTweetQuarters* are not statistically significant. However, once the sample is restricted to verified accounts only, in columns (3) and (4), coefficient estimates for *FinNewsTweetQuarters*

Table 6: Tweeting and Information Diffusion

This table shows the cross-sectional relationship between momentum in monthly stock returns and the consistency of tweeting about earnings news. The dependent variable is *Momentum*; in columns (1) and (3) is calculated using excess returns relative to 90 day T-bills and in columns (2) and (4) using Fama-French three factor excess returns. Columns (3) and (4) are estimated using the subsample of firms with verified Twitter accounts. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. See Appendix table A5 for variable definitions. The regression is estimated using OLS with robust standard errors. Standard errors are provided in parentheses beneath the coefficient estimates. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

	All Firms		*Verified Twitter Firms*	
	(1)	(2)	(3)	(4)
Fin News Tweet Quarters	-0.018	-0.002	-0.041***	-0.040**
	(0.011)	(0.013)	(0.016)	(0.020)
BM	0.022	-0.007	0.038*	0.032
	(0.015)	(0.034)	(0.020)	(0.041)
Analysts	0.001	-0.001	0.001	0.000
	(0.001)	(0.001)	(0.001)	(0.001)
Size	0.000	0.004	0.004	0.007
	(0.004)	(0.005)	(0.005)	(0.008)
Institutional Ownership	0.010	0.014	0.013	0.003
	(0.032)	(0.036)	(0.049)	(0.066)
Twitter Followers	0.000	-0.000	0.000	-0.000
	(0.000)	(0.000)	(0.000)	(0.000)
Constant	-0.090*	-0.120**	-0.176**	-0.190**
	(0.053)	(0.050)	(0.073)	(0.094)
R^2	0.054	0.022	0.064	0.047
Observations	1,064	848	443	356

are -0.041 and -0.040 and are statistically significant at the 1% and 5% levels, respectively. This negative relationship suggests firms with verified accounts that tweet about earnings news more consistently have less momentum in returns. This result suggests firms may be able to increase the speed of information diffusion to investors by tweeting about earnings news. This result is consistent with media's role to disseminate information quickly.

F. Content analysis of tweets

My analysis confirms that financial news shared on Twitter has a significant impact on the market value of equities in some situations. Specifically, I find that firms with negative earnings surprises tend to have higher daily announcement returns when they tweet about their earnings announcement.

The equity returns patterns that I document are consistent with the model in section 3. These results may also be consistent with models featuring hidden information and adverse selection. To distinguish among these possible models, I provide data from a formal comparison of tweets across positive and negative earnings surprise days.

Table 7 shows the results from a feature extraction exercise. In columns (4)–(6) I consider all tweets about financial news in a three-day window around earnings announcement events. In columns (1)–(3) I consider the remaining tweets, those unrelated to financial news, in the same three-day window. The results are aggregate across firms and quarters. Quarters are divided into two groups, positive and negative earnings surprise quarters relative to the analyst consensus forecast.

The features in Table 7 are comprehensive and enable me to study information content, readability, sentiment, and attention. I note several interesting patterns. First, there is significant heterogeneity between tweets posted on positive versus negative earnings surprise days. Financial news tweets receive less attention on days with negative earnings surprises than on those with positive surprises (as measured by *Likes* and *Retweets*). This suggests that on average the diffusion of good news will be faster than that of bad news. The sentiment of financial news tweets is also less positive on days with negative earnings surprises than on those with positive surprises (as measured by *Positive sentiment* and *Compound sentiment*). This result is not surprising, but it suggests that the relatively positive effects of tweeting on bad news days cannot be explained by sentiment alone. In fact, the sentiment of tweets seems to be in line with the news itself, indicating that managers are not using sentiment strategically. The measures of sentiment I use are based on a VADER (Valence Aware Dictionary for sEntiment Reasoning) model, which is sensitive to both polarity (positive/negative) and intensity (strength) of emotion and is often used in performing sentiment analysis on social media data (available at https://pypi.org/project/vaderSentiment/). *Positive* and *Negative sentiment* are the proportions of text that fall into these categories. In contrast, *Compound sentiment* is a metric that calculates the sum of all the lexicon ratings and normalizes them between -1 and 1.

Financial news tweets tend to be easier to read on days with negative earnings surprises than on days with positive ones, as measured by *Readability Index*, *Word count*, *Characters*, and *Difficult words*. These metrics show that financial news tweets about negative surprises tend to be shorter, to contain fewer difficult words, and to be overall easier to understand. The *Readability Index* is a consensus score based on the most common methods for calculating the grade level of a text (available at https://pypi.org/project/textstat/). A score of 9.2, for instance, means that a ninth grader would typically be able to read the text.

Table 8 shows the 60 most common unigrams and bigrams in the corpus of financial news tweets. The tweets are divided into two groups, those posted on positive earnings surprise days and those posted on negative surprise days. The total frequency of appearance of each word in positive and negative earnings surprise tweets and the average frequency per tweet are reported. I note several interesting differences between financial news tweets on positive and negative announcement days.

First, financial news tweets are significantly less likely to mention “EPS” (earnings per share) on days with negative earnings surprises than positive surprise days. This result is particularly interesting because of the definition of negative news, in which actual EPS is compared to the market’s expected EPS. Also, EPS is one of the most important numbers released during quarterly and annual announcements, attracting analysts’ attention and media coverage. This result suggests that firms use discretion when announcing financial news on social media and are less likely to disseminate an unfavorable metric.

Second, financial news tweets are twice as likely to mention “dividends” on days with negative earnings surprises than positive surprise days. Like earnings per share, dividends are closely watched by investors and communicate the financial well-being of a firm. This result suggests that firms may use Twitter to republicize “good news” on days when their earnings results are poor.

Table 9 shows the 60 most common unigrams and bigrams in the remaining tweets in my sample. There is significant heterogeneity between tweets posted on positive earnings surprise days and negative surprise days. In comparison to Table 8, the average frequency of terms per tweet is much lower, this is to be expected, as these tweets span a wider range of topics than the financial news tweets.

Table 7: Features of Tweets

This table provides a mean comparisons of statistics from tweets when firms have a negative or positive earnings surprise. I test for differences in means using a t-test. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

	Additional Tweets			**Financial News Tweets**		
	Pos Surprise	*Neg Surprise*	Δ	*Pos Surprise*	*Neg Surprise*	Δ
	(1)	(2)	(3)	(4)	(5)	(6)
Likes	40.166	37.248	-2.918	7.23	4.16	-3.070***
Retweets	22.549	17.197	-5.352	3.809	2.597	-1.212***
URL	0.753	0.728	-0.024***	0.848	0.864	0.017*
Picture	0.401	0.385	-0.016***	0.145	0.118	-0.027***
User tags	0.442	0.437	-0.004	0.132	0.116	-0.016
Hashtags	0.896	0.919	0.023**	0.482	0.405	-0.077***
Percentages	0.034	0.032	-0.002	0.152	0.127	-0.025**
Dollar amounts	0.029	0.035	0.007***	0.157	0.153	-0.004
Negative sentiment	0.028	0.026	-0.002***	0.011	0.012	0.001
Positive sentiment	0.151	0.158	0.007***	0.074	0.066	-0.008**
Compound sentiment	0.248	0.267	0.019***	0.150	0.130	-0.021**
Readability Index	8.835	8.645	-0.189***	9.208	8.992	-0.216**
Word count	13.789	13.818	0.028	14.923	14.406	-0.516***
Characters	89.182	88.843	-0.340	97.072	94.090	-2.981**
Difficult words	3.751	3.579	-0.173***	4.468	4.330	-0.138**
Syllables per word	1.665	1.650	-0.015***	1.667	1.675	0.008

Table 8: Financial news tweets: Term frequency

This table provides a mean comparisons of statistics from tweets when firms have a negative or positive earnings surprise. I test for differences in means using a t-test. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

		Positive Earnings Surprise		Negative Earnings Surprise		
Order	Common Terms	Frequency	Per tweet	Frequency	Per tweet	Δ
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	earnings	4343	0.47	1147	0.45	-0.020
2	results	3831	0.41	1156	0.45	0.039***
3	quarter	2957	0.38	923	0.43	0.046***
4	reports	1980	0.21	640	0.25	0.037***
5	year	1074	0.18	328	0.20	0.023*
6	financial	1275	0.14	382	0.15	0.012
7	financial results	950	0.10	292	0.11	0.012
8	today	746	0.09	204	0.10	0.007
9	webcast	797	0.09	227	0.09	0.002
10	sales	773	0.08	187	0.07	-0.010
11	second quarter	601	0.08	170	0.08	0.000
12	growth	716	0.08	139	0.05	-0.022***
13	ceo	713	0.08	167	0.07	-0.011
14	revenue	710	0.08	150	0.06	-0.018**
15	eps	698	0.08	119	0.05	-0.029***
16	conference	648	0.07	131	0.05	-0.019***
17	announces	645	0.07	267	0.10	0.035***
18	release	642	0.07	175	0.07	-0.001
19	fiscal	591	0.06	181	0.07	0.007
20	net	586	0.06	140	0.06	-0.008
21	fourth quarter	472	0.06	195	0.09	0.030***
22	strong	572	0.06	97	0.04	-0.023***
23	share	551	0.06	160	0.06	0.002
24	live	501	0.06	141	0.06	0.001
25	record	476	0.05	68	0.03	-0.026***
26	quarter results	798	0.05	259	0.05	-0.003
27	billion	471	0.05	103	0.04	-0.010
28	tomorrow	432	0.05	99	0.04	-0.007
29	join	1469	0.05	526	0.04	-0.001
30	million	409	0.04	153	0.06	0.016**
31	income	381	0.04	123	0.05	0.006
32	listen	385	0.04	129	0.05	0.009
33	reported	358	0.04	113	0.04	0.005
34	earnings conference	352	0.04	86	0.03	-0.004
35	guidance	347	0.04	91	0.04	-0.002
36	operating	343	0.04	90	0.04	-0.001
37	revenues	320	0.04	63	0.03	-0.010*
38	quarter year	196	0.04	85	0.05	0.015***
39	read	320	0.03	103	0.04	0.006
40	quarter earnings	387	0.03	126	0.04	0.005
41	adjusted	260	0.03	70	0.03	-0.001
42	earnings results	268	0.03	89	0.03	0.006
43	net income	255	0.03	74	0.03	0.001
44	cfo	249	0.03	71	0.03	0.001
45	performance	247	0.03	36	0.01	-0.012***
46	details	242	0.03	76	0.03	0.004
47	learn	241	0.03	44	0.02	-0.009*
48	press release	237	0.03	75	0.03	0.004
49	grew	236	0.03	34	0.01	-0.012***
50	increased	235	0.03	55	0.02	-0.004
51	announced	233	0.03	82	0.03	0.007
52	diluted	231	0.03	33	0.01	-0.012***
53	business	213	0.02	62	0.03	0.002
54	fiscal year	195	0.02	65	0.03	0.005
55	cash	191	0.02	69	0.03	0.006
56	dividend	191	0.02	95	0.04	0.016***
57	fy	639	0.01	152	0.01	0.000
58	quarter financial	297	0.01	84	0.01	0.000
59	fullyear	345	0.00	99	0.00	0.000
60	nongaap	282	0.00	42	0.00	0.000

Table 9: Additional tweets: Term frequency

This table provides a mean comparisons of statistics from tweets when firms have a negative or positive earnings surprise. I test for differences in means using a t-test. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

		Positive Earnings Surprise		Negative Earnings Surprise		
Order	Common Terms	Frequency	Per tweet	Frequency	Per tweet	Δ
(1)	(2)	(3)	(4)	(5)	(6)	(7)
1	new	4888	0.06	1540	0.06	-0.004*
2	learn	3226	0.04	973	0.04	-0.004**
3	today	2462	0.04	899	0.04	0.005**
4	help	2321	0.03	758	0.03	-0.001
5	day	1991	0.03	736	0.03	0.003**
6	check	1871	0.02	672	0.03	0.002
7	ceo	1785	0.02	404	0.02	-0.007***
8	make	1575	0.02	508	0.02	-0.001
9	business	1572	0.02	492	0.02	-0.001
10	time	1572	0.02	543	0.02	0.001
11	know	1543	0.02	515	0.02	0.000
12	great	1518	0.02	563	0.02	0.002*
13	data	1457	0.02	446	0.02	-0.002
14	join	1453	0.02	514	0.02	0.001
15	read	1394	0.02	480	0.02	0.001
16	need	1369	0.02	394	0.02	-0.003**
17	growth	1308	0.02	327	0.01	-0.004***
18	best	1241	0.02	437	0.02	0.001
19	video	1189	0.02	358	0.01	-0.002
20	like	1185	0.02	416	0.02	0.001
21	watch	1162	0.02	321	0.01	-0.002**
22	booth	1127	0.02	447	0.02	0.003**
23	team	1120	0.02	361	0.01	0.000
24	thanks	1093	0.01	369	0.01	0.000
25	look	1082	0.01	414	0.02	0.002*
26	work	1051	0.01	345	0.01	-0.001
27	week	1050	0.02	426	0.02	0.004***
28	want	1034	0.01	345	0.01	0.000
29	tech	1004	0.01	275	0.01	-0.003***
30	live	1000	0.01	283	0.01	-0.002*
31	win	1000	0.01	394	0.02	0.002**
32	digital	996	0.01	314	0.01	-0.001
33	use	967	0.01	337	0.01	0.000
34	home	953	0.01	371	0.02	0.002*
35	visit	909	0.01	295	0.01	0.000
36	future	905	0.01	278	0.01	-0.001
37	tips	891	0.01	380	0.02	0.003***
38	technology	881	0.01	239	0.01	-0.002**
39	share	873	0.01	286	0.01	0.000
40	global	856	0.01	288	0.01	0.000
41	customers	846	0.01	281	0.01	0.000
42	good	846	0.01	336	0.01	0.002*
43	market	843	0.01	335	0.01	0.002**
44	job	840	0.01	314	0.01	0.001
45	world	828	0.01	253	0.01	-0.001
46	happy	827	0.01	320	0.01	0.002*
47	years	818	0.01	242	0.01	-0.001
48	love	813	0.01	356	0.01	0.003***
49	latest	807	0.01	234	0.01	-0.001
50	free	795	0.01	301	0.01	0.001
51	blog	794	0.01	265	0.01	0.000
52	cloud	792	0.01	291	0.01	0.001
53	support	780	0.01	235	0.01	-0.001
54	energy	779	0.01	337	0.01	0.004***
55	health	775	0.01	228	0.01	-0.001
56	security	774	0.01	240	0.01	-0.001
57	did	766	0.01	260	0.01	0.000
58	looking	766	0.01	273	0.01	0.000
59	way	766	0.01	263	0.01	0.000
60	people	762	0.01	227	0.01	-0.001

5. Robustness

In this section I report the results of various robustness tests confirming the results in this paper. I show that the relationship between tweeting and announcement returns is robust to method used to determine unexpected earnings is defined, to the sample selection, and to additional fixed effect specifications.

A potential concern is that I overlook important information by using the dummy variable *Negative Surprise* rather than the continuous variable *SUE*. Both variables measure the surprise of the earnings announcement relative to the market's expectations. *Negative Surprise* is equal to one when a firm announces earnings below the analyst consensus forecast, and zero otherwise. *SUE* is the firm's actual earnings minus the analyst consensus forecast of earnings, standardized by the standard deviation of analyst forecasts.

Figure 4 shows the distribution of standardized unexpected earnings across the sample. In Table 10, I replace *Negative Surprise* with *SUE* and re-estimate equation (2). The coefficient on the interaction term *SUE * Fin News Tweet Impact* is negative and significant across specifications. This suggests that the price response to financial news tweets depends on how positive or negative the quarterly announcement is. These results are consistent with the main analysis.

In Table 11, I replace *Negative Surprise* with three bins that capture the distribution of *SUE*. I split the sample into quartiles by *SUE* and define the variables *Quartile 1*, *Quartile 2*, and *Quartile 3*. Each of these variables is equal to 1 when a firm's *SUE* is in that quartile of the distribution, and zero otherwise. *Quartile 4* is the omitted (reference) group. The coefficient on the interaction term *Q1* × *Fin News Tweet Impact* is positive and significant at 1% across specifications. The coefficients on the interaction terms *Q2* × *Fin News Tweet Impact* and *Q3* × *Fin News Tweet Impact* are also positive and significant; however, when the dummy coefficient is added to the interaction term, the full effects are insignificant. At the bottom of the table I show the results of an F-test. These results are consistent with the main analysis.

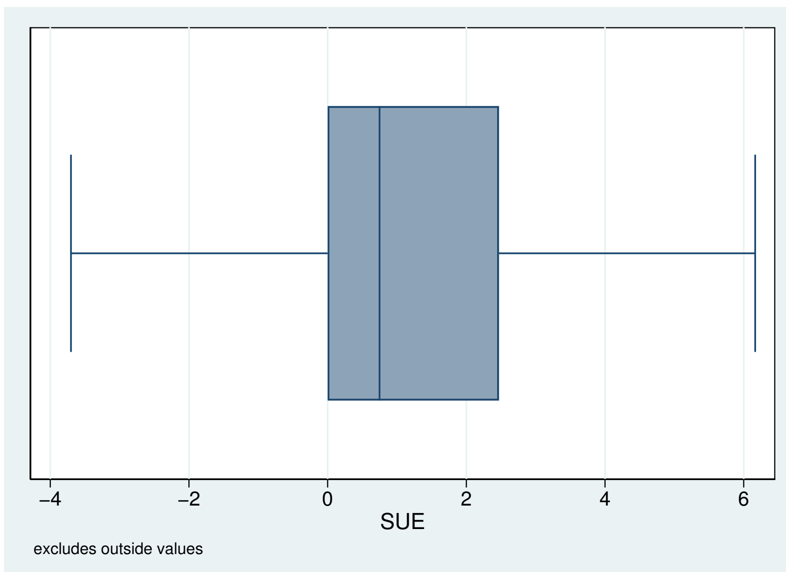


Figure 4: Box-plot of standardized unexpected earnings This figure depicts the distribution of standardized unexpected earnings (SUE) in my sample.

Table 10: Tweeting and Announcement Returns

This table shows the relationship between cumulative abnormal returns, $CAR_{[-1,1]}$, and a firm's tweeting behavior. The dummy variable *Negative Surprise* is replaced by the continuous variable *SUE*, standardized unexpected earnings. In columns (1) and (2) *FinNewsTweetImpact* is measured as *FinNewsTweets*, in columns (3) and (4) as *FinNewsTweets*Retweets*. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. Standard errors are clustered at the firm level and provided in parentheses beneath the coefficient estimates. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

	CAR			
	(1)	(2)	(3)	(4)
SUE	0.703***	0.767***	0.703***	0.737***
	(0.028)	(0.034)	(0.028)	(0.031)
Fin News Tweet Impact	0.283*	0.414***	0.076	0.236***
	(0.149)	(0.153)	(0.075)	(0.082)
*SUE * Fin News Tweet Impact*		-0.123***		-0.076***
		(0.039)		(0.016)
Additional Tweet Impact	0.048	0.057	0.055	0.047
	(0.083)	(0.085)	(0.083)	(0.035)
Residual ESV (Ryans)	-0.421	-0.409	-0.421	-0.429
	(0.288)	(0.289)	(0.288)	(0.289)
Size	-2.610***	-2.606***	-2.615***	-2.626***
	(0.422)	(0.442)	(0.422)	(0.421)
Loss	-1.826***	-1.733***	-1.837***	-1.802***
	(0.305)	(0.335)	(0.305)	(0.304)
BM	7.096***	7.384***	7.076***	7.084***
	(0.756)	(0.767)	(0.753)	(0.752)
Q4	0.336	0.153	0.333	0.325
	(0.208)	(0.225)	(0.208)	(0.207)
Firm FE	Yes	Yes	Yes	Yes
Quarter-year FE	Yes	Yes	Yes	Yes
F-test statistic: $\beta 2 + \beta 3 = 0$		4.117**		4.226**
F-test p-value		0.043		0.040
No. of firms	1067	1054	1067	1067
Adjusted R^2	0.127	0.143	0.126	0.128
Observations	14,222	13,170	14,222	14,222

Table 11: Tweeting and Announcement Returns: SUE Quartiles

This table shows the relationship between cumulative abnormal returns, $CAR_{[-1,1]}$, and a firm's tweeting behavior. The variable *Negative Surprise* is replaced by three variables, *Quartile 1*, *Quartile 2*, and *Quartile 3*. Each of these variables is a equal to 1 when a firm's standardized unexpected earnings, *SUE*, is in that quartile of the distribution and equal to zero otherwise. In columns (1) and (2) *FinNewsTweetImpact* is measured as *FinNewsTweets*, in columns (3) and (4) as *FinNewsTweets*Retweets*. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. Standard errors are clustered at the firm level and provided in parentheses beneath the coefficient estimates. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

	CAR			
	(1)	(2)	(3)	(4)
Fin News Tweet Impact	0.291**	-0.302	0.073	0.069
	(0.146)	(0.204)	(0.073)	(0.159)
Quartile 1 (SUE)	-7.100***	-7.672***	-7.105***	-7.398***
	(0.241)	(0.288)	(0.241)	(0.257)
Quartile 2 (SUE)	-3.653***	-4.049***	-3.656***	-3.822***
	(0.206)	(0.255)	(0.207)	(0.222)
Quartile 3 (SUE)	-2.354***	-2.605***	-2.358***	-2.458***
	(0.173)	(0.220)	(0.173)	(0.188)
*Q1 * FinNewsTweetImpact*		1.268***		0.607***
		(0.291)		(0.110)
*Q2 * FinNewsTweetImpact*		0.824***		0.294***
		(0.242)		(0.100)
*Q3 * FinNewsTweetImpact*		0.491**		0.157*
		(0.196)		(0.086)
Additional Tweet Impact	0.030	0.027	0.053	0.050
	(0.082)	(0.083)	(0.034)	(0.034)
Residual ESV (Ryans)	-0.435	-0.430	-0.448	-0.439
	(0.287)	(0.288)	(0.287)	(0.287)
Size	-2.355***	-2.362***	-2.374***	-2.373***
	(0.402)	(0.403)	(0.402)	(0.403)
Loss	-1.896***	-1.874***	-1.903***	-1.860***
	(0.309)	(0.308)	(0.309)	(0.309)
BM	7.070***	7.067***	7.062***	7.091***
	(0.756)	(0.756)	(0.753)	(0.755)
Fourth Quarter	0.331	0.324	0.328	0.316
	(0.203)	(0.204)	(0.203)	(0.203)
Firm FE	Yes	Yes	Yes	Yes
Quarter-year FE	Yes	Yes	Yes	Yes
F-test statistic Q1: $\beta 1 + \beta 5 = 0$		16.56***		14.99***
F-test statistic Q2: $\beta 1 + \beta 6 = 0$		6.63		5.05
F-test statistic Q3: $\beta 1 + \beta 7 = 0$		1.16		2.22
No. of firms	1067	1067	1067	1067
Adjusted R^2	0.138	0.140	0.138	0.140
Observations	14,222	14,222	14,222	14,222

In Table 12, I show that the results are robust to a pooled OLS estimation and a rich set of fixed effects. Fixed effects help to control for unobservable determinants of tweeting: quarter-year fixed effects for macro factors, and firm-year (firm) fixed effects for time-varying (time-invariant) firm characteristics. The results are generally consistent across specifications; however, the within-group estimates tend to be higher and more significant than the pooled OLS estimates.

Earnings announcement-specific characteristics can also bias the estimates. Firms may be more likely to disclose bad news on Friday than on Monday–Thursday (DellaVigna and Pollet, 2009). To control for the variation of announcements on different days, I use day-of-week fixed effects. To control for observable announcement-specific characteristics, I include the variables $SUE_{i,t}$, $Q4_{i,t}$, and $Loss_{i,t}$.

One concern is that some of the Twitter accounts I manually collected could be erroneous or fake accounts. To control for this potential problem I limit the sample to those firms with verified Twitter accounts. The verified feature on Twitter is a signal to the public that an account of public interest is authentic. Of the 1,215 accounts in my sample, 489 are verified. Table 13 shows that the subsample of verified firms yields results similar to those in section 4.B.

The relationship between tweeting and momentum in returns is robust to different momentum proxies. In Table 6 I calculate *Momentum* using both excess returns relative to 90-day T-bills and Fama-French excess returns. In Table 14, I measure momentum in three alternative ways. Following Hong et al. (2000) I use the serial correlation coefficient of six-month excess returns (relative to 90-day T-bills). I also calculate *momentum (AC)* using cumulative 3-month excess returns rather than monthly returns. *Momentum* is calculated in column (1) using 3-month excess returns relative to 90-day T-bills and in column (2) using Fama-French three-factor excess returns.

Table 12: Various Fixed Effects: Tweeting and Announcement Returns

This table shows the relationship between cumulative abnormal returns, $CAR_{[-1,1]}$, and firms' tweeting behaviors. In columns (1) and (2) *FinNewsTweetImpact* is measured as *FinNewsTweets*, in column (3) as *FinNewsTweets*Followers*, and in column (4) as *FinNewsTweets*Retweets*. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. Standard errors are clustered at the firm level and provided in parentheses beneath the coefficient estimates. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

	CAR				
	(1)	(2)	(3)	(4)	(5)
Negative Surprise	-3.632***	-3.622***	-3.771***	-3.968***	-3.970***
	(0.205)	(0.206)	(0.214)	(0.243)	(0.243)
Fin News Tweet Impact	-0.193**	-0.192**	0.086	0.181	0.177
	(0.091)	(0.092)	(0.152)	(0.218)	(0.218)
*Neg Surp*Fin News Tweet Impact*	0.747***	0.727***	0.778***	0.879***	0.875***
	(0.212)	(0.211)	(0.229)	(0.251)	(0.251)
Additional Tweet Impact	-0.055	-0.067	0.044	0.091	0.106
	(0.051)	(0.052)	(0.081)	(0.110)	(0.110)
Residual ESV (Ryans)	-0.409	-0.392	-0.408	-0.599**	-0.590**
	(0.284)	(0.289)	(0.284)	(0.292)	(0.292)
SUE	0.417***	0.419***	0.464***	0.483***	0.483***
	(0.026)	(0.026)	(0.028)	(0.031)	(0.031)
Size	-0.236***	-0.228***	-2.557***	-4.818***	-4.835***
	(0.039)	(0.039)	(0.361)	(0.941)	(0.939)
Loss	-1.206***	-1.178***	-1.637***	-2.085***	-2.085***
	(0.242)	(0.243)	(0.301)	(0.346)	(0.346)
BM	1.687***	1.662***	6.873***	20.462***	20.492***
	(0.212)	(0.213)	(0.710)	(1.723)	(1.720)
Q4	0.494***	0.340	0.492***	0.397*	0.389*
	(0.144)	(0.207)	(0.143)	(0.210)	(0.210)
Quarter-year FE	No	Yes	No	Yes	Yes
Firm FE	No	No	Yes	No	No
Firm FE	No	No	No	Yes	Yes
Weekday FE	No	No	No	No	Yes
F-test statistic: $\beta 2 + \beta 3 = 0$	10.462	9.879	14.151	14.506	14.312
F-test p-value	0.001	0.002	0.000	0.000	0.000
No. of firms	1067	1067	1067	1046	1046
Adjusted R^2	0.130	0.130	0.156	0.193	0.193
Observations	14,223	14,222	14,223	13,838	13,838

Table 13: Verified Firms: Tweeting and Announcement Returns

This table shows the relationship between cumulative abnormal returns, $CAR_{[-1,1]}$, and firms' tweeting behaviors for the subsample of firms with verified Twitter accounts. In columns (1) and (2) *FinNewsTweet-Impact* is measured as *FinNewsTweets*, in column (3) as *FinNewsTweets*Followers*, and in column (4) as *FinNewsTweets*Retweets*. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. Standard errors are clustered at the firm level and provided in parentheses beneath the coefficient estimates. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

	CAR			
	(1)	(2)	(3)	(4)
Negative Surprise	-3.255***	-3.619***	-3.255***	-3.530***
	(0.251)	(0.288)	(0.251)	(0.271)
Fin News Tweet Impact	0.016	-0.166	0.013	-0.059
	(0.169)	(0.179)	(0.080)	(0.082)
*Neg Surp*Fin News Tweet Impact*		0.854***		0.456***
		(0.290)		(0.114)
Additional Tweet Impact	-0.019	-0.015	-0.019	0.061
	(0.106)	(0.107)	(0.106)	(0.038)
SUE	0.501***	0.499***	0.500***	0.500***
	(0.040)	(0.040)	(0.040)	(0.040)
Residual ESV (Ryans)	-0.533	-0.536	-0.533	-0.547
	(0.398)	(0.399)	(0.398)	(0.400)
Size	-2.591***	-2.563***	-2.592***	-2.605***
	(0.604)	(0.607)	(0.604)	(0.608)
Loss	-1.774***	-1.751***	-1.772***	-1.738***
	(0.422)	(0.423)	(0.423)	(0.423)
BM	6.536***	6.562***	6.538***	6.583***
	(1.199)	(1.196)	(1.201)	(1.204)
Q4	0.597**	0.582**	0.597**	0.588**
	(0.270)	(0.270)	(0.270)	(0.269)
Firm FE	Yes	Yes	Yes	Yes
Quarter-year FE	Yes	Yes	Yes	Yes
F-test statistic: $\beta2 + \beta3 = 0$		5.867		9.581
F-test p-value		0.016		0.002
No. of firms	463	463	463	463
Adjusted R^2	0.159	0.160	0.159	0.161
Observations	7,357	7,357	7,357	7,357

Table 14: Tweeting and short-term continuation in returns

This table shows the cross-sectional relationship between short-term continuation in returns and the consistency of tweeting about earnings news. The dependent variable is measured in three ways. *AC* is measured as the correlation between the series $ExRet_{i,[t,t+2]}$ and the lagged series $ExRet_{i,[t-12,t-2]}$, where $ExRet_{i,[t,t+2]}$ is the cumulative 3-month excess return of firm *i*. In column (1) *AC* is calculated using excess returns relative to 90 day T-bills and in column (2) using Fama-French three factor excess returns. In column (3) *SCC* is the serial correlation of six-month excess returns (relative to 90 day T-bills). The sample is restricted to firms with verified Twitter accounts. To mitigate the influence of outliers, all continuous variables are winsorized at the 1st and 99th percentiles. See Appendix A for variable definitions. The regression is estimated using OLS with robust standard errors. Standard errors are provided in parentheses beneath the coefficient estimates. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

	AC		*SCC*
	(1)	(2)	(3)
Fin News Tweet Quarters	-0.048*	-0.057**	-0.017*
	(0.025)	(0.028)	(0.009)
BM	0.024	0.039	0.023***
	(0.028)	(0.044)	(0.005)
Analysts	-0.000	-0.000	0.001*
	(0.001)	(0.002)	(0.000)
Size	0.016**	0.001	-0.003
	(0.007)	(0.010)	(0.002)
Institutional Ownership	0.003	-0.067	-0.018
	(0.084)	(0.100)	(0.024)
Twitter Followers	0.000*	-0.000	0.000
	(0.000)	(0.000)	(0.000)
Constant	-0.284***	-0.116	0.799***
	(0.099)	(0.121)	(0.027)
R^2	0.031	0.022	0.036
Observations	442	352	476

6. Conclusion

This paper studies the link between firms' voluntary disclosure strategies on social media and their equity returns using novel data of over 7 million tweets posted by S&P 1500 firms. Since regulators do not require disclosures on social media, this has created a unique empirical setting where disclosures vary over time and across firms. This paper takes advantage of this variation to help identify the impact of social media disclosures on equity returns while carefully considering managers' incentives.

One of the main empirical challenges that this paper addresses is separating the effect of the disclosure decision from the effect of the information being disclosed. In particular, I focus on a specific class of tweets that disseminate financial results after mandatory quarterly disclosure events. This setting allows me to study tweets that capture the firm's disclosure strategy but do not convey additional news to the market. Furthermore, by focusing on earnings announcements, I can use analysts' forecasts to control for the market's expectations.

I use a model to shed light on the mechanisms through which strategic voluntary disclosure impacts investors' expectations and, ultimately, the price of firms' equity. The model provides three key empirical implications, which I test using my dataset. First, the model suggests that firms will disclose their successes and withhold their failures. Second, the model predicts that stock prices will rise more for firms that follow a strategic disclosure policy following a relatively poor earnings announcement. The intuition is that the marginal benefit of strategically disseminating information on social media is higher for firms that are less likely to have good news to disclose. Finally, the model predicts that stock prices will rise more for firms with higher levels of retail investor ownership.

I document three main results consistent with models of strategic disclosure. First, I characterize firms' strategic use of Twitter. I find that firms tweet more after good news and strategically use tweets' tone and information content. Second, I find that firms with negative earnings surprises have higher announcement returns when they tweet about financial news, suggesting that firms can use social media to bolster their stock prices during periods of poor performance. Finally, I provide evidence that the disclosures on social media matter more for retail investors, consistent with social media being a primary information source for investors with a high cost of information acquisition and processing.

The findings of this study are of importance to regulators, investors, and firms. Social media is a new disclosure channel that has gained an outreach as relevant as traditional information intermediaries, such as business press, newswire services, and financial analysts. Nevertheless, the choice to disclose information on social media channels has been left to the managers' discretion. Despite the SEC's attempt to promote full and fair disclosures, the information a firm discloses on social media often reflects managers' strategic decisions.

References

Amir, E. and Livnat, J. (2005). The economic consequences of (not) issuing preliminary earnings announcement, *Technical report*, Working paper, New York University.

Bartov, E., Faurel, L. and Mohanram, P. (2017). Can twitter help predict firm-level earnings and stock returns?, *The Accounting Review* .

Bartov, E., Givoly, D. and Hayn, C. (2002). The rewards to meeting or beating earnings expectations, *Journal of Accounting and Economics* **33**(2): 173 – 204.

Bhagwat, V. and Burch, T. (2016). Pump it up? tweeting to manage investor attention to earnings news.

Blankespoor, E., Miller, G. S. and White, H. D. (2014). The role of dissemination in market liquidity: Evidence from firms' use of twitter™, *The Accounting Review* **89**(1): 79–112.

Carhart, M. M. (1997). On persistence in mutual fund performance, *The Journal of finance* **52**(1): 57–82.

Cohen, L. and Frazzini, A. (2008). Economic links and predictable returns, *The Journal of Finance* **63**(4): 1977–2011.

Davis, A. K., Ge, W., Matsumoto, D. and Zhang, J. L. (2015). The effect of manager-specific optimism on the tone of earnings conference calls, *Review of Accounting Studies* **20**(2): 639–673.

DellaVigna, S. and Pollet, J. M. (2009). Investor inattention and friday earnings announcements, *The Journal of Finance* **64**(2): 709–749.

Dye, R. A. (1985). Disclosure of nonproprietary information, *Journal of accounting research* pp. 123–145.

Engelberg, J. E. and Parsons, C. A. (2011). The causal impact of media in financial markets, *The Journal of Finance* **66**(1): 67–97.

Fang, L. and Peress, J. (2009). Media coverage and the cross-section of stock returns, *The Journal of Finance* **64**(5): 2023–2052.

Fedyk, A. (2018). Front page news: The effect of news positioning on financial markets.

Goto, S., Watanabe, M. and Xu, Y. (2009). Strategic disclosure and stock returns: Theory and evidence from us cross-listing, *The Review of Financial Studies* **22**(4): 1585–1620.

Grossman, S. J. (1981). The informational role of warranties and private disclosure about product quality, *The Journal of Law and Economics* **24**(3): 461–483.

Grossman, S. J. and Hart, O. D. (1980). Takeover bids, the free-rider problem, and the theory of the corporation, *The Bell Journal of Economics* pp. 42–64.

Hirshleifer, D. and Teoh, S. H. (2003). Limited attention, information disclosure, and financial reporting, *Journal of accounting and economics* **36**(1-3): 337–386.

Hong, H., Lim, T. and Stein, J. C. (2000). Bad news travels slowly: Size, analyst coverage, and the profitability of momentum strategies, *The Journal of Finance* **55**(1): 265–295.

Hong, H. and Stein, J. C. (1999). A unified theory of underreaction, momentum trading, and overreaction in asset markets, *The Journal of finance* **54**(6): 2143–2184.

Huberman, G. and Regev, T. (2001). Contagious speculation and a cure for cancer: A nonevent that made stock prices soar, *The Journal of Finance* **56**(1): 387–396.

Jegadeesh, N. and Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency, *The Journal of finance* **48**(1): 65–91.

Jung, M. J., Naughton, J. P., Tahoun, A. and Wang, C. (2018). Do firms strategically disseminate? evidence from corporate use of social media, *The Accounting Review* **93**(4): 225–252.

Kothari, S. P., Shu, S. and Wysocki, P. D. (2009). Do managers withhold bad news?, *Journal of Accounting Research* **47**(1): 241–276.

Loughran, T. and McDonald, B. (2017). The use of edgar filings by investors, *Journal of Behavioral Finance* **18**(2): 231–248.

Matsumoto, D. A. (2002). Management's incentives to avoid negative earnings surprises, *The accounting review* **77**(3): 483–514.

Milgrom, P. R. (1981). Good news and bad news: Representation theorems and applications, *The Bell Journal of Economics* pp. 380–391.

Peng, L. and Xiong, W. (2006). Investor attention, overconfidence and category learning, *Journal of Financial Economics* **80**(3): 563–602.

Rogers, J. L., Van Buskirk, A. and Zechman, S. L. (2011). Disclosure tone and shareholder litigation, *The Accounting Review* **86**(6): 2155–2183.

Ryans, J. (2017). Using the edgar log file data set, *Available at SSRN 2913612* .

SEC (2013). Sec says social media ok for company announcements if investors are alerted, *Press Release* pp. April 2, 2013.

Shin, H. S. (2003). Disclosures and asset returns, *Econometrica* **71**(1): 105–133.

Skinner, D. J. and Sloan, R. G. (2002). Earnings surprises, growth expectations, and stock returns or don't let an earnings torpedo sink your portfolio, *Review of accounting studies* **7**(2): 289–312.

Tetlock, P. C. (2007). Giving content to investor sentiment: The role of media in the stock market, *The Journal of finance* **62**(3): 1139–1168.

Appendix

Comparison of firms with and without Twitter accounts

To determine which types of firms have a corporate Twitter account, I compare a broad set of firm characteristic along with valuation, liquidity, profitability, and financial soundness metrics by estimating the following regression:

$$\begin{aligned} TwitterDummy_{i,t} = \alpha &+ \beta_1 Size_{i,t} + \beta_2 StockMarketIndex_{i_t} + \beta_3 Valuation_{i,t} \\ &+ \beta_4 Profitability_{i,t} + \beta_5 FinancialSoundness_{i,t} \\ &+ \beta_6 IndustryChars + \beta_7 OtherRatios. \end{aligned} \tag{5}$$

In equation (5), the dependent variable, $TwitterDummy$, is a dummy equal to 1 if the firm has an active Twitter account as of October 2017. $Size$ is the natural logarithm of *Total Assets*. $StockMarketIndex$ is composed of two dummy variables, *S&P 500* and *S&P 600* which indicate whether the firm was listed in the respective index. $Valuation$ is composed of the *Book to Market* and *Price to Operating Earnings* ratios. $Profitability$ is composed of the *Gross Profit Margin* and *Gross Profit to Total Assets* ratios. $FinancialSoundness$ is composed of *Capitalization*, *Cash Balance to Total Liabilities*, *Long Term Debt to Total Liabilities*, *Operating CF to Current Liabilities*, and *Asset Turnover*. $IndustryChars$ is composed of the *Research and Development over Sales*, *Advertising Expenses over Sales*, *Labor Expenses over Sales*, an indicator variable if a firm is in a manufacturing industry (*Manufacturing*), and an indicator variable if a firm is in a business-to-consumer traded industry (*B2C Traded Industry*). Finally, $OtherRatios$ is composed of *Accruals over Average Assets* and *Institutional Ownership as a Percentage of Shares Outstanding*.

In Table A1, equation (5) is estimated with standard errors clustered by industry.[9] In column (1) a probit model is estimated, while in columns (2) and (3) OLS and OLS with industry fixed effects models are estimated, respectively.

[9]The three-digit North American Industry Classification System (NAICS) code is used for clustering.

On average, larger firms and, incremental to size, firms belonging to the S&P 500 index have a higher probability of having a Twitter account. This result suggests that having a Twitter account is not a substitute for overall visibility but rather a complement to it.

Firms with a lower book-to-market ratio also have a higher probability of having a Twitter account. Technology companies and other companies in industries that have fewer physical assets tend to have a low book-to-market ratio. However, this result holds when including industry fixed effects (column (3)), and therefore it appears that firms with relatively higher valuations than their industry peers are more likely to have a Twitter account.

Table A1: Firms With and Without Twitter

This table shows the relationship between the likelihood of having a corporate Twitter account and various firm characteristics. The dependent variable is a dummy variable equal to one if a firm has a Twitter account during my sample period, and equal to zero otherwise. In columns (1) the regression is estimated using a probit model, in columns (2) and (3) using an OLS model. Standard errors are clustered by industry and are provided in parentheses beneath the coefficient estimates. See Appendix A for variable definitions. *, **, and *** indicates significance at the 10%, 5%, and 1%, respectively.

	Twitter Dummy		
	(1)	(2)	(3)
Size (log of Total Assets)	0.206***	0.060***	0.064***
	(0.036)	(0.010)	(0.011)
S&P 500 (Large Cap)	0.215**	0.064*	0.071*
	(0.105)	(0.033)	(0.036)
S&P 600 (Small Cap)	0.032	-0.003	-0.005
	(0.082)	(0.030)	(0.030)
Book/Market	-0.209***	-0.067***	-0.035
	(0.070)	(0.019)	(0.022)
Price/Operating Earnings	0.000*	0.000*	0.000
	(0.000)	(0.000)	(0.000)
Gross Profit Margin	0.314**	0.066*	0.036
	(0.147)	(0.036)	(0.035)
Gross Profit/Total Assets	0.537***	0.171***	0.189***
	(0.206)	(0.057)	(0.057)
Capitalization Ratio	0.121	0.038	0.019
	(0.262)	(0.080)	(0.081)
Cash Balance/Total Liabilities	0.021	0.006	0.004
	(0.036)	(0.012)	(0.011)
Long-term Debt/Total Liabilities	-0.478	-0.163	-0.126
	(0.292)	(0.098)	(0.098)
Operating CF/Current Liabilities	-0.122**	-0.032**	-0.025**
	(0.058)	(0.014)	(0.011)
Asset Turnover	0.071	0.016	0.012
	(0.058)	(0.017)	(0.019)
Research and Development/Sales	0.453**	0.101*	0.061
	(0.204)	(0.051)	(0.049)
Advertising Expenses/Sales	-0.908	-0.063	-0.096
	(1.171)	(0.156)	(0.165)
Labor Expenses/Sales	0.954**	0.104*	0.056
	(0.485)	(0.053)	(0.038)
Accruals/Average Assets	0.170	0.020	0.080
	(0.371)	(0.122)	(0.128)
Institutional Ownership % Shrs Out	-0.131	-0.005	-0.043
	(0.211)	(0.070)	(0.073)
Manufacturing	-0.119	-0.050	
	(0.103)	(0.034)	
B2C Traded Service	0.153	0.038	
	(0.161)	(0.046)	
Industry FE	No	No	Yes
No. of clusters	74	74	78
Adjusted (Pseudo) R^2	(0.078)	0.085	0.159
Observations	29,584	29,584	29,774

Summary statistics

Table A2: Distribution of Tweets by Calendar Quarter

This table presents the frequency distributions of tweets and observations by calendar quarter. My sample encompasses 7,132,461 tweets posted by S&P1500 firms with active Twitter accounts as of October 2017. The sample represents represents 16,844 firm-quarters. The number of firm quarter observations increases over the sample. This pattern is to be expected because some Twitter users in the sample were not active at the start of the sample period.

Calendar Quarter	Tweets		Observations	
	N	%	N	%
2014Q1	422,377	5.4%	950	5.0%
2014Q2	428,965	5.5%	968	5.1%
2014Q3	460,574	5.9%	984	5.2%
2014Q4	501,935	6.5%	996	5.2%
2015Q1	463,611	6.0%	1,023	5.4%
2015Q2	481,138	6.2%	1,032	5.4%
2015Q3	489,600	6.3%	1,060	5.5%
2015Q4	570,359	7.3%	1,070	5.6%
2016Q1	433,123	5.6%	1,058	5.5%
2016Q2	435,102	5.6%	1,071	5.6%
2016Q3	420,817	5.4%	1,070	5.6%
2016Q4	454,652	5.9%	1,089	5.7%
2017Q1	428,492	5.5%	1,103	5.8%
2017Q2	378,364	4.9%	1,110	5.8%
2017Q3	384,522	5.0%	1,128	5.9%
2017Q4	378,830	4.9%	1,132	5.9%
All	7,132,461	100.0%	16,844	100.0%

Table A3: Tweet Characteristics

This table presents descriptive statistics related to Twitter users (firms), firm-quarters, and individual tweets. Firms have a mean (median) or 162,642 (6,352) followers. The average date firms joined Twitter was in November 2010. Firm-quarters have a mean (median) of 178 (81) tweets and 1.24 (1) tweets about earnings news. Tweets have a mean (median) of 79 (86) characters, 8 (0) retweets, and 16 (0) likes.

Variable	Mean	Std. Dev	P01	Q1	Median	Q3	P99
Per Twitter User (N = 1,215)							
Number of Followers	162,642	1,383,262	73	1,473	6,352	29,200	2,300,412
Number of Friends	2,438	9,814	0	194	557	1,535	35,626
Date Joined Twitter	Nov2010	-	Jun2007	Apr2009	Jan2007	Jan2012	May2017
Per Firm Quarter (N = 13,350)							
Tweet Count	178.00	430.00	0.00	22.00	81.00	209.00	1438.00
Quarter with Tweets	92%	0.26	0.00	1.00	1.00	1.00	1.00
Earnings Tweet Count	1.24	3.48	0.00	0.00	0.00	1.00	13
Quarter with Earnings Tweets	35%	0.48	0.00	0.00	0.00	1.00	1.00
Per Tweet (N = 7,132,461)							
Number of Characters	79.00	50.00	16.00	18.00	86.00	119.00	217.00
Number of Retweets	8.00	276.00	0.00	0.00	0.00	1.00	149.00
Number of Likes	16.00	753.00	0.00	0.00	0.00	2.00	264.00

Table A4: Summary Statistics

This table reports summary statistics for firm-quarter observations used to estimate (1) and (2), and firm-month observations to estimate (3). The sample period is from Q1 2014 to Q4 2017. See Appendix table A5 for variable definitions.

	Mean	SD	P05	Med	P95
CAR[-1, 1]	0.16	7.51	-11.58	0.19	11.85
Positive Surprise	0.72	0.45	0.00	1.00	1.00
Earnings Tweet Count [-1, 1]	0.76	1.93	0.00	0.00	4.00
Non-earnings Tweet Count [-1, 1]	2.37	1.44	0.00	2.56	4.49
*Earnings Tweet Count*Followers [-1, 1]*	7.61	20.97	0.00	0.00	40.31
*Non-earnings Count*Followers [-1, 1]*	24.05	17.44	0.00	22.89	53.78
*Earnings Tweet Count*Retweets [-1, 1]*	0.74	2.62	0.00	0.00	3.85
*Non-earnings Count*Retweets [-1, 1]*	3.16	3.90	0.00	1.88	11.21
$\mid SUE \mid$	2.58	3.06	0.00	1.60	9.03
Size	8.60	1.76	5.90	8.49	11.74
Loss	0.12	0.33	0.00	0.00	1.00
BM	0.48	0.35	0.08	0.39	1.14
Analysts	2.60	0.62	1.61	2.65	3.50
Q4	0.20	0.40	0.00	0.00	1.00
SUE (forecast sd)	1.47	3.61	-3.23	0.97	7.78
SUE (price)	0.00	0.01	-0.00	0.00	0.01
SUE (book equity)	0.00	0.01	-0.01	0.00	0.02
Earnings Tweet Count [-30, -1]	0.31	1.08	0.00	0.00	2.00
Non-Earnings Tweet Count [-30, -1]	3.15	1.60	0.00	3.40	5.35
Institutional Own.	0.85	0.14	0.60	0.86	1.05
Momentum	-10.07	12.75	-30.38	-10.84	11.21
Earnings Tweet Quarters	0.34	0.38	0.00	0.18	1.00
*Earnings Tweet Quarters*Followers*	3.14	3.68	0.00	1.48	10.63
Verified	0.42	0.49	0.00	0.00	1.00

Variable definitions

Table A5: Variable Descriptions and Data Sources

Variable	Description	Data Source
Twitter Variables		
Earnings Tweet Impact	Measured in one of two ways depending on model: (1) IHS transformation of *Financial News Tweet Count*, (2) IHS transformation of *Financial News Tweet Count* * IHS transformation of *Financial News Retweets*	Twitter
Financial News Tweet Count	Number of earnings related tweets during the windows [-1, 1] around the quarterly earnings announcement date	Twitter
Additional Tweet Impact	Measured in one of two ways depending on model: (1) IHS transformation of *Additional Tweet Count*, (2) IHS transformation of *Additional Tweet Count* * IHS transformation of *Additional Retweets*	Twitter
Additional Tweet Count	IHS transformation of total number of tweets minus the number financial news tweets during the windows [-1, 1] around the quarterly earnings announcement date	Twitter
Earnings Tweet Quarters	Proportion of quarters a firm tweets about financial news over the sample period, January 2014 to December 2017	Twitter
Twitter Verified	Indicator variable equal to one is a firm's Twitter account is verified by Twitter. When an account is verified by Twitter a blue check-mark appears next to the account name to signal the authenticity of that account.	Twitter
Earnings Announcement Variables		
CAR[-1, 1]	Carhart's cumulative abnormal return in the three day window [-1, 1] around the earnings announcement date	CRSP
Negative Surprise	Indicator variable equal to one if the firm's $SUE < 0$, and equal to zero otherwise.	IBES

SUE	The firm's actual EPS minus the consensus analyst forecast EPS, standardized by the standard deviation of analysts' consensus forecasts, by price per share of stock at the end of the quarter, or by the book value of equity per share at the end of the previous quarter. Consensus analyst forecast is measured as the median latest analyst forecast in the 90 days prior to the earnings announcement.	IBES
Firm Variables		
Size	Log of total assets (Compustat atq).	Compustat
BM	Book to market value (Compustat ceqq/mkvaltq).	Compustat
Loss	Indicator variable set to 1 if the firm reports a quarterly loss (Compustat $niq < 0$).	Compustat
Analyst	Natural log of one plus the average number of analysts following a given firm during the 90 days prior to the earnings announcement.	IBES
Q4	Indicator variable equal to one if the quarterly earnings announcement is in the fourth fiscal quarter of the year	Compustat
Institutional Ownership %	Total institutional ownership as a percentage of shares outstanding.	Thomson Reuters 13-f
Autocovariance Variables		
Autocovariance	Correlation between the series $EzRet_{[t,t]}$ and the lagged series $ExRet_{[t-12,t-2]}$, where $ExRet_{[t,t]}$ is the monthly excess return of a firm. Excess returns are calculated relative to 90 day T-bills or Fama-French three factor excess returns.	CRSP
EDGAR Log File Variables		
ESSV	Daily EDGAR Search Volume from the SEC's web server log file. Before calculating ESV the log files are filtered to remove downloads by computer programs following the procedure developed by Ryans (2017) or Loughran and McDonald (2017).	SEC

Financial news key words and phrases

To detect financial news tweets I use a classification scheme based on the dictionary of keywords and phrases below; each tweet is considered earnings news if it contains two or more of the terms found in the dictionary.

Financial news unigrams, bigrams, and trigrams: *announce, announces, cash flow, conference call, continuing operations, declare, declares, dividend, dividends, earnings, earnings call, earnings release, eps, financial position, financial results, fiscal, full year, gaap, growth, income, net sales, press release, profit, releases, results, revenue, sales, $"ticker of firm", 1q, 2q, 3q, 4q, q1, q2, q3, q4, qtr1, qrt2, qrt3, qrt4, 1st quarter, 2nd quarter, 3rd quarter, 4th quarter, first quarter, second quarter, third quarter, fourth quarter, quarter, qtr, qoq, fy13, fy14, fy15, fy16, fy17, fy18, fy2013, fy2014, fy2015, fy2016, fy2017, fy2018, year-over-year, year over year, yoy*

Examples of financial news tweets identified with text classifier

Figure 5: Tweeting around earnings announcements. This figure depicts examples of financial news tweets in my sample. Earnings announcement keywords and phrases are outlined in red. Each tweet in my sample is considered financial news if it contains two or more of the terms highlighted in red.

Theoretical model

The following equations are re-stated from Goto et al. (2009).

The firm value at date 0 is given by

$$V_0 = [\psi u + (1-\psi)d]^N, \tag{1}$$

where $\psi \equiv ru^{-\alpha}/[ru^{-\alpha} + (1-r)d^{-\alpha}]$.

When managers follow a strategic disclosure strategy (i.e., manager reports the observed number of successes, s, and zero failures at date 1) the firm value at date 1 is given by

$$V_1(s) = [\pi u + (1-\pi)d]^{N-s}, \tag{2}$$

where $\pi \equiv qu^{-\alpha}/[qu^{-\alpha} + (1-q)d^{-\alpha}]$ and $q \equiv (r - r\theta)/(1-\theta r)$.

Therefore the expected first-period return under strategic disclosure is given by

$$E[R_1(s)] = \sum_{s=0}^{N} h(s)R_1(s) = [r\theta\gamma_0 + (1-r\theta)\gamma_1]^N, \tag{3}$$

where $h(s) = \binom{N}{s}(r\theta)^s(1-r\theta)^{N-s}$ is the unconditional probability of the manager announcing s successes at date 1, $\gamma_0 \equiv u/[\psi u + (1-\psi)d] > 1$, and $\gamma_1 \equiv [\pi u + (1-\pi)d]/[\psi u + (1-\psi)d] < 1$.

When managers follow a full disclosure strategy (i.e., manager reports the observed number of successes, s, and failures, f, at date 1) the firm value at date 1 is given by

$$V_1(s,f) = [\psi u + (1-\pi)d]^{N-s-f}u^s d^f, \tag{4}$$

Therefore the expected first-period return under full disclosure is given by

$$E[R_1(s,f)] = [r\theta\gamma_3 + (1-r\theta)]^N > 1, \tag{5}$$

where $\gamma_3 \equiv [ru + (1-r)d]/[\psi u + (1-\psi)d] > 1$.

Using the above equations I estimate the difference between expected first period returns when managers use a strategic disclosure strategy and when managers use a full disclosure strategy. Figure 6 shows the difference in expected returns as a function of r. When the function is monotonically decreasing this implies that the expected increase in return under strategic disclosure strategy is higher for firms which announce a negative earnings surprise at date 0 than for firms that announce positive earnings surprise. The vertical dotted line at 0.73 represents a reasonable value of r estimated in my sample, which falls in the area where the function is downward sloping.

Figure 7 shows the difference in expected returns as a function of θ. When θ increases managers are more likely to observe the outcomes of their business dimensions at $t = 1$ and asymmetric information between managers and investors is higher. Therefore, θ can be thought of as a measure of relative information asymmetry. Under reasonable parametric assumptions the model predicts that the jump in expected returns associated with using a strategic disclosure strategy is increasing in θ.

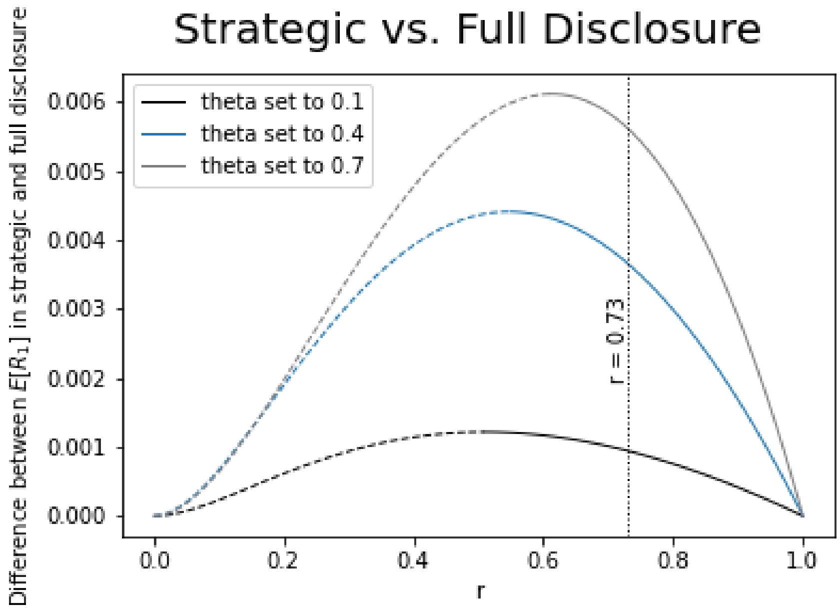


Figure 6: Difference in expected first period returns (as a function of r). This figure depicts the difference between expected first period returns when managers use a strategic disclosure strategy and when managers use a full disclosure strategy as a funciton of r, the probability a project succeeds. The other parameter values are set at $N = 100$, $u = 1.001$, $d = .99$, and $\alpha = 3$.

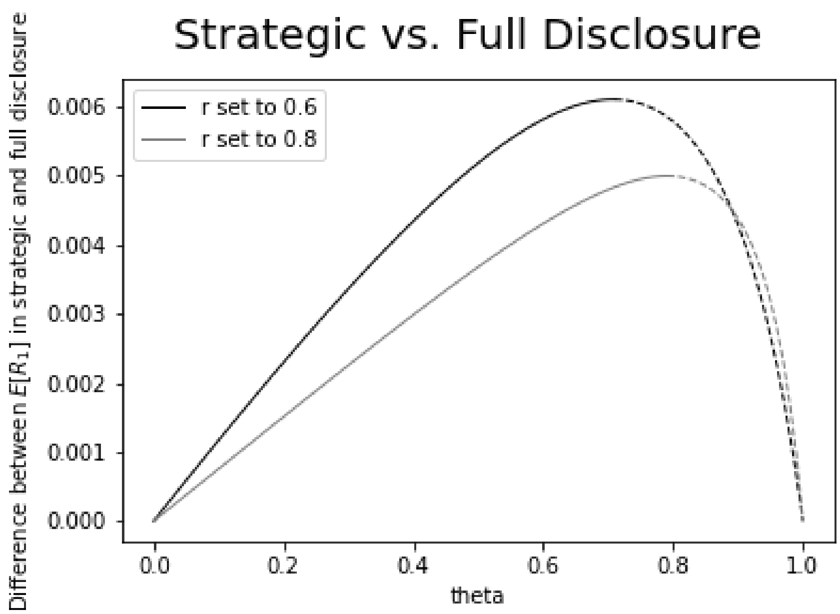


Figure 7: Difference in expected first period returns (as a function of θ). This figure depicts the difference between expected first period returns when managers use a strategic disclosure strategy and when managers use a full disclosure strategy as a funciton of r, the probability a project succeeds. The other parameter values are set at $N = 100$, $u = 1.001$, $d = .99$, and $\alpha = 3$.

Chapter 2. When the Markets Get CO.V.I.D: COntagion, Viruses, and Information Diffusion

1 Introduction

COVID19 has manifested itself as a very aggressive and fast epidemic that—at the time of the first draft of this paper—brought major economic countries to their knees.[1] Given the fast-increasing contagion curve of COVID19 and its global scale, this epidemic event is challenging common economic policy interventions and depressing the global value of our assets, i.e., the wealth of millions of households all over the world.

Given that severe virus-related crises are expected to become more frequent, we find it relevant to use COVID19-related data to ask the following broad questions about financial market reactions to viral contagion risk. First, what is the average impact of medical announcements on financial returns? Equivalently, is the diffusion of this information enhancing wealth or adding risk? Second, what is the market price of risk of news related to global contagion dynamics? Third, can local contagion conditions help us predict expected returns?

Last but not least, can we use social media activity to measure production and diffusion of information about epidemic risk? This question is important for at least two reasons. First, fast epidemic outbreaks tend to get investors off guard and hence real-time indexes based on social media news may function as a useful predictive tool. Second, the estimation of multidimensional models requires many observations that we may gather by using high-frequency data, as opposed to waiting for daily medical bulletins.

[1] Our first draft is dated 3/23/2020. To assess the severity of COVID19, see the March 11, 2020 WHO Director-General's opening remarks (`https://www.who.int/dg/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19---11-march-2020`).

In this study, we address these questions by quantifying the exposure of major financial markets to news shocks about global contagion risk accounting for local epidemic conditions. For a wide cross section of countries, we construct a novel data set comprising (i) medical announcements related to COVID19; and (ii) high-frequency data on epidemic news diffused through Twitter. Across several classes of financial assets and currencies, we provide novel empirical evidence about financial dynamics (i) around epidemic announcements, (ii) at a daily frequency, and (iii) at an intra-daily frequency. Formal estimations based on both contagion data and social media activity about COVID19 confirm that the market price of epidemic risk is very significant. We conclude that prudential policies aimed at mitigating either global contagion or local diffusion may be extremely valuable.

Current results in detail. An important contribution of our work is the collection of a novel dataset on the COVID19 pandemic that includes (i) a very large set of official announcements on medical conditions (about 13,000 announcements), and (ii) news diffused on Twitter in real-time by major newspapers (based on more than 753,000 tweets). We identify major newspapers for a large cross section of major countries in the spirit of Baker et al. (2016). In contrast to Baker et al. (2016), we do not analyze articles, rather we track news published on Twitter in real time, so that we can produce high frequency data when needed.

More specifically, we track tweets posted by major newspapers with key words such as 'coronavirus' and 'covid19'. For each newspaper, we identify the location of its headquarters so that we can identify its specific time-zone. As a result, we gather thousands of tweets for a large cross section of countries that we can aggregate at different frequencies and across regions.

Given this data set, we document several important facts about news diffusion. First, both Twitter-based news diffusion (measured by number of tweets) and attention (measured by number of retweets) spike upon contagion-related announcements. Second and more broadly, the diffusion of information increases substantially in each country in our data set as soon as that country goes into an epidemic state.[2] Third, our measured increase in information diffusion is particularly pronounced precisely during the hours in which financial markets are open. All of these empirical facts suggest that tracking Twitter-diffused news can be a reliable way to characterize the information set of investors at a high frequency.

Turning our attention to financial dynamics, we look at equity returns around announcements, that is, in a ± 60 minute window. We find that cumulative equity returns have no clear pattern before the announcement, as they tend to be relatively flat and indistinguishable from zero. In the post-announcement time window, instead, cumulated returns jump upward. This result holds also when we focus only on announcements of bad news and it is present also in countries with relatively high contagion levels. Furthermore, the positive average effect of medical announcements on equities is present both upon local and foreign announcements.

We note that this time behavior of returns is not present in the pre-epidemic state and is quite different from that documented in Lucca and Moench (2015). Lucca and Moench (2015) shows a slow and persistent accumulation of positive returns before monetary policy announcements. In our case, instead, the increase in the cumulative returns at the announcement is consistent with the Ai and Bansal (2018) model. When the representative investor cares about the timing of resolution of uncertainty, prices jump upward when uncertainty is resolved along the information cycle and then they start to decline. Future research should study whether similar patterns could be due to overreaction (Bordalo et al. 2020).

Furthermore, we conduct the same analysis looking at the government bond market. The response of bonds is less severe than that observed in equities. In a ± 60-minute window around the announcement, there is no significant adjustment in bond returns among advanced economies (AE). At first, this result may look surprising as bonds may be in higher demand since considered safer assets. Hence one may expect to find an average appreciation. On the other hand, one may expect that default concerns generate a simultaneous downward pressure on bond prices. Since we

[2] We identify the beginning of the epidemic state with the day in which the number of confirmed COVID19 cases becomes greater than or equal to 100.

find a modest link between COVID19 news and default concerns as measured by CDS quotes, we speculate that this result is mainly driven by monetary policy.

Among emerging economies (EE), in contrast, bond prices experience a positive sudden increase around announcements, but it is less relevant than that for equities. By no-arbitrage, this observation suggests that cash-flow uncertainty is an important determinant of the market fluctuations observed during the COVID19 crisis. This high-frequency result is consistent with the results documented by Gormsen and Koijen (2020) looking at dividend futures.

We also look at equity market trading volume around announcement times and document that it exhibits an upward adjustement upon the announcement time and then a slow reversal. We show that this pattern is less severe for AE than for EE. When we look at bid-ask spreads of sovereign bonds, we find an immediate reduction upon announcements for AE and a delayed one for EE. The magnitude of the decline in the bid-ask spread is comparable across AE and EE. Taken together, these patters suggest that investors are active with safer assets both in AE and EE.

In the last step of our analysis, we group countries into three portfolios on a daily basis according to their relative number of COVID19 cases. We do this separately for advanced and EE. The H (L) portfolio comprises the equity returns of the top (bottom) countries in terms of COVID19 contagion cases. We then estimate a no-arbitrage based model in which we allow for time-varying betas ($\beta_{i,t}$) with respect to global contagion risk. Specifically we allow equity returns to respond to global viral contagion news according to the relative share of official COVID19 cases associated to each portfolio. Global contagion risk is measured either by innovations in the growth rate of global COVID19 contagion cases or by innovations in the tone of our COVID19-related tweets.

This model can potentially capture many of the features of equity returns that we document in our descriptive analysis. First, this model captures predictability through contagion-based time-varying betas. Second, this specification has the potential to capture higher negative skewness for countries that go through more severe contagion paths. Consider the case of portfolio H comprising countries receiving a sequence of relatively more severe contagion news. This portfolio will have greater exposure to adverse news ($|\beta_{H,t}|$ increases) as the relative contagion share of the portfolio grows. As the relative contagion share starts to flatten out and eventually decline, the sensitivity of this portfolio to good news is reduced ($|\beta_{H,t}|$ shrinks), meaning that returns will be less sensitive to positive news and hence the right tail of their distribution will not be very long.

Third, this model accounts for heterogeneous exposure to global contagion news and hence it enables us to identify the market price of risk of this global contagion component. Across all of our specifications, the market price of contagion risk is both statistically significant and extremely high. Equities are more exposed to risk than bonds. Both within advanced and EE, heterogeneous exposure to contagion risk is substantial and as a result an equity-based HML-COVID strategy bears a high risk premium. An HML-COVID strategy that goes long in bonds of countries with a larger share of cases and short a smaller share of cases, instead, provides an insurance premium. This means that in countries very exposed to contagion risk, bonds tend to become safer. We find that this result is particularly sizable among EE.

These results conform well with the data on weekly international investment flows. Countries with lower (higher) contagion levels are expected to experience equity inflows (outflows). Expected inflows are stronger in AE than in EE. In contrast, when looking at bonds, these findings are almost absent in AE, and reversed in EE, meaning that in high-covid emerging economies the flows going toward government bonds increase. This is consistent with the idea that bonds are perceived as safer assets in EE.

In the last step of our analysis, we run intra-day regressions taking advantage of our high-frequency Twitter-based risk measure. We focus on European countries whose markets are open simultaneously, namely, ITA, ESP, UK, FRA, DEU, CHE, and SWE. Every day, we group them into three portfolios according to their relative number of COVID19 cases measured in the previous 24 hours. The H (L) portfolio comprises the equity returns of the top-2 (bottom-2) countries for COVID19 contagion cases.

Our novel high-frequency estimation confirms our main findings: policies related to prevention and containment of contagion could be very valuable not only in terms of lives saved but also in terms of global wealth. These results hold also after controlling for the market and for changes in equity volatility. Our results have been very stable over time and can be explored at `https://sites.google.com/view/when-markets-get-covid/`, a website that we use for the visualization of our data.[3]

Related literature. Due to its relevance, the COVID19 crisis has spurred a lot of contemporaneous research (Goldstein et al. 2021). Macroeconomic studies are focusing on both the aggregate and distributional dynamic implications of the epidemic crisis (Hagedorn and Mitman 2020; Coibion et al. 2020; Eichenbaum et al. 2020; Fornaro and Wolf 2020; Chiou and Tucker 2020; Barrot et al. 2020; Alon et al. 2020; Glover et al. 2020; Corsetti et al. 2020; Caballero and Simsek 2020; Coven and Gupta 2020; Hensvik et al. 2020).

Other analyses assess policy concerns (Acemoglu et al. 2020; Alvarez et al. 2020; Jones et al. 2020; Bahaj and Reis 2020; Elgin et al. 2020; Faria-e Castro and Louis 2020; Krueger et al. 2020; Farboodi et al. 2020). Correia et al. (2020) and Barro et al. (2020) provide evidence using data from the 1918-Flu epidemic. We differ from these studies for our strong attention to asset prices and COVID19-driven risk.

Other studies at the intersection of macroeconomics and econometrics focus on forecasting the diffusion of both contagion cases and COVID19-implied economic activity disruptions (Favero 2020; Ichino et al. 2020; Atkeson 2020; Atkeson 2020; Ma et al. 2020; Ludvigson et al. 2020). We focus on both the cross sectional and time series implications for asset prices across different asset classes.

An important strand of the literature focuses on the measurement of both COVID19-induced uncertainty and firm-level risk exposure by utilizing textual analysis and surveys (Baker et al. 2020b; Hassan et al. 2020; Bartik et al. 2020). Giglio et al. (2020) use a survey to study investor expectations over different horizons. Lewis et al. (2020) provide a novel weekly measure of economic

[3]Our updates are schedule in October, January, and May.

activity using several labor market-based timeseries. We focus on high-frequency data, Twitter-based news diffusion, epidemic announcements, and country-level asset price dynamics. Our study adds viral contagion risk considerations to the findings of Pelger (2020).

Gerding et al. (2020) look at equity market dynamics and link the epidemic risk exposure to country-level fiscal capacity. Augustin et al. (2021) looks at CDS. Bonaccolto et al. (2019) focus on currency union break up risk due to COVID19. Papanikolaou and Schmidt (2020) look at the financial implications of industry-level job disruption due to COVID19. Albuquerque et al. (2020) focus on the performance of firms with high environmental and social ratings during the COVID19 outbreak. They do not study announcements and they do not assess the market price of viral contagion risk. Ramelli and Wagner (2020) study equity returns across firms accounting for international trade, financial strength, and investor attention. They use both Google search volume and conference calls as a measure of attention, whereas we use high-frequency data on retweets of tweets issued by news provider. Pástor and Vorsatz (2020), Baker et al. (2020a), Bretscher et al. (2020a), and Kaniel and Wang (2020) study the impact of COVID19 on financial markets. We provide novel evidence about both (i) market reactions around contagion-related announcement times, and (ii) the market price of contagion risk at high frequency.

Schoenfeld (2020) examines buy-and-hold returns for many assets and finds that managers systematically underestimate their exposure to COVID19. Cororaton and Rosen (2020) look at the characteristics of firms participating to the US Paycheck Protection Program. Acharya and Steffen (2020) study firm-loan-level data to study the implications for liquidity. Carletti et al. (2020) look at Italian firms. Alfaro et al. (2020) focus on the link between aggregate equity market returns and unanticipated changes in predicted infections during the SARS and COVID19 pandemics. Bretscher et al. (2020b) look at the supply channel of uncertainty shocks. Hartley and Rebucci (2020) and Sinagl (2020) look at monetary policy announcements and cash-flow risk, respectively. Cox et al. (2020) confirm the relevance of monetary policy estimating a dynamic asset pricing model. We differ in our attention to medical announcements; our social media-based measures of information diffusion and attention; and our high frequency analysis. Our work complements the evidence in

Gormsen and Koijen (2020) and Gormsen et al. (2021) who extract relevant information about expectations and risk premia from derivatives.

Within the literature that studies news coverage reaction to news, our manuscript is methodologically related to the work of, among others, Bianchi et al. (2021), Hassan et al. (2019), Manela and Moreira (2017), Garmaise et al. (2021), Tetlock (2007), Calomiris and Mamaysky (2019), Israelsen et al. (2021), Cookson et al. (2021), Bybee et al. (2020) and Engle et al. (2020).

2 Medical Announcements

In this section, we illustrate key features of our novel data set comprising thousands of COVID19-related announcements across twenty one countries. We then show our main results. Specifically, we document that: (i) equity markets on average appreciate upon announcements, and especially so in emerging economies (henceforth EEs); (ii) bond returns are insensitive to announcements in advanced economies (henceforth AEs), but appreciate to some extent in EEs; (iii) across both AEs and EEs, trade becomes more active after medical announcements.

2.1 Data Collection

We treat the release of each medical bulletin as an announcement. The same applies to travel limitations and lock down policies related to COVID19. We note that we have manually tracked these policy interventions on a daily basis and hence we have constructed a novel dataset important to study real-time high frequency reactions of financial markets to epidemic risk.

Since in our sample we have also witnessed important announcements related to both monetary and fiscal policy interventions, we complement the medical announcements with major policy-related announcements as well. We note that medical announcements in our sample period are much more prominent than policy-related announcements as they represent nearly 86% of all of

Table 1. **Summary Statistics for Announcements**

Country	**No. Announcements**	**Governments & Central Banks**	**Med. Bulletins & Lockdowns**
Argentina	424	30.66%	69.34%
Australia	617	0.00%	100.00%
Brazil	800	38.00%	62.00%
Canada	569	0.00%	100.00%
Chile	726	29.20%	70.80%
China	426	0.00%	100.00%
Colombia	827	29.14%	70.86%
France	381	2.10%	97.90%
Germany	288	2.78%	97.22%
Hong Kong	1,149	0.00%	100.00%
India	592	0.68%	99.32%
Italy	468	4.91%	95.09%
Japan	122	69.67%	30.33%
Korea	558	0.18%	99.82%
Mexico	1,624	35.47%	64.53%
New Zealand	426	0.00%	100.00%
Spain	447	2.01%	97.99%
Sweden	341	0.00%	100.00%
Switzerland	449	0.45%	99.55%
UK	523	1.91%	98.09%
USA	1,147	1.48%	98.52%
Total	12,904	12.63%	87.37%

Notes: This table shows summary statistics for COVID19-related announcements that we collect for a large cross section of countries. Our real-time data range from 1/1/2020 to the date of this manuscript. For each country, we report the total number of announcements, the fraction related to either medical bulletins or lock-down measures, as well as the fraction of other announcements issued by governments and central banks about fiscal and monetary policy, respectively.

the announcements collected. Our data collection is very comprehensive, as documented in table 1, and it comprises more than 10,000 medical announcements. An example of a COVID19-related announcement follows:

> **2020-03-14 15:35:00; Vice President @Mike_Pence and members of the Coronavirus Task Force will hold a press briefing at 12:00 p.m. ET. Watch LIVE:** `http://45.wh.gov/RtVRmD`

In this case, we set the time of the announcement at 12:00 p.m. ET. To clarify further our methodology, we also give an example of an announcement related to a monetary policy intervention

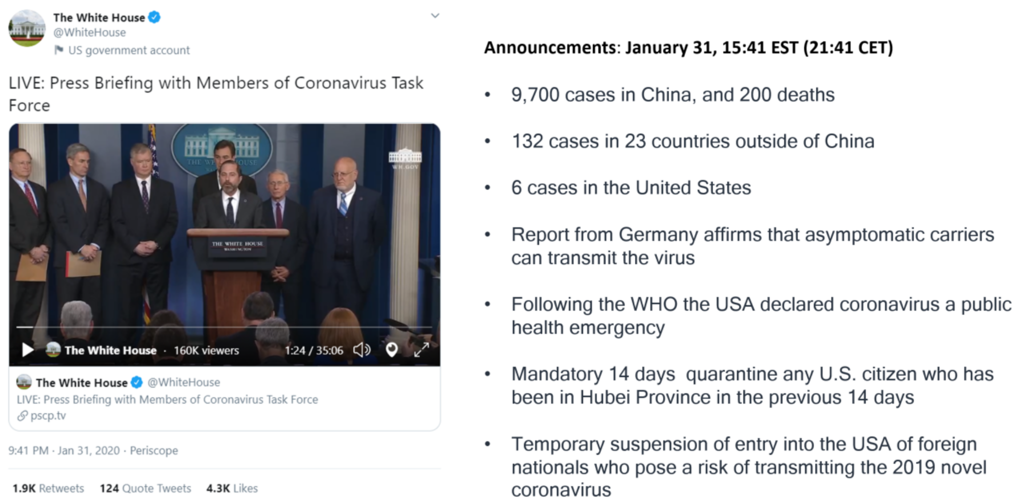


Fig. 1. Announcement Time from Twitter.

Notes: This figure shows a tweet about one of the first COVID-related announcements in the US. The tweet time stamp enables us to identify the effective timing of the announcement. On the right hand side of this figure, we summarize the topics discussed during the briefing.

in response to COVID19:

> **2020-03-18 23:05:00; FT Breaking News; ECB to launch €750bn bond-buying programme.**

In this case, the time of the announcement is 11:05 p.m. CET.

We 'hand-collect' these announcements in several ways. First of all, for each country we look for official press statements publicly available on the webpage of the local Ministry of Health (MoH). If the press statement does not have an official time stamp, we look for it on the official Twitter account of the MoH or other related government entities (for example, the Twitter account of the Prime Minister). If this second attempt fails as well, we look at the Twitter accounts of major local newspapers and focus on news about medical reports. These steps, which we repeat multiple times during each week, are sufficient to identify the effective time of each announcements in our data set relevant for financial investors.

As an example, in figure 1 we report our record of the first scheduled Coronavirus Task Force Press briefing. In contrast to the following White House press meetings, this briefing took place earlier, at 3:40 p.m. EST. This example demonstrates two important aspects of our dataset construction: (i) it accounts for meetings scheduled at not-recurrent times; and (ii) it captures purely COVID-related news.[4]

2.2 Announcements and Financial Markets

Pre- and post-epidemic samples. In what follows, we study the financial dynamics around medical announcement times. In order to isolate the dynamics related solely to medical announcements, we plot the differential behavior of our variable of interest with respect to normal times, i.e., pre-epidemic times. In each country, we define the beginning of the epidemic period as the day in which the country experienced an official number of contagion cases greater than or equal to 100. Given this threshold, China is the first country in our sample to go in the epidemic phase, whereas New Zealand is last.

The pre-epidemic sample starts for all countries on October 1st 2019 so that the pre-epidemic period comprises at least four months of data. This subsample is long enough to run meaningful comparisons with the post-pandemic subsample. More specifically, consider, for example, an announcement on a Friday at 3:40 p.m. EST. We compare the reaction of our financial variables around this announcement to their behavior at the same time across all of the Fridays comprised in our pre-epidemic sample.

Pre- and post-announcement behavior. We run a high-frequency analysis around announcement times. In what follows, we estimate the following regression at the minute-level:

$$Z_t = (c_{pre} + c_{t>t^*}) + (\alpha_{pre} + \alpha_{t>t^*}) \cdot t + (\beta_{pre} + \beta_{t>t^*}) \cdot t^2, \quad t \in [t^* \pm K] \tag{1}$$

[4]Our dataset enables researchers to easily identify each specifc announcement and hence look for the content discussed in each one of the events that we detect.

where t^* is the time of the announcement, K is equal to 60 minutes; and Z_t is the differential behavior of our variable of interest across the pre- and post-epidemic sample. This specification is a quadratic function of time that includes dummy variables to account for post-announcement jumps in both the level and the slope. We test the null assumption that there is no difference post-announcement, $H_0 : c_{t>t^*} = \alpha_{t>t^*} = \beta_{t>t^*} = 0$, and if we fail to reject the null we depict the resulting smooth quadratic fit. Standard errors are always HAC-adjusted.

Information Diffusion. Our novel social media-based data set enables us to measure the diffusion of information at a very high frequency. For each announcement in our data set, we compile all COVID-related tweets issued in a $\pm$60-minute window around announcement time by major newspapers in each country. We provide a detailed description of our data collection procedure in the next section. For the sake of statistical power, we aggregate all of these tweets across all of our countries and we call the resulting aggregate 'World'.

In the left panel of figure 2, we show per-country per-minute average number of tweets around announcement times during epidemic periods in excess of the same average measured in the pre-epidemic samples (dots). This procedure enables us to capture news diffusion patterns specific to the epidemic period. The right panel refers to retweets, that is, our measure of attention to the news.

Formal tests reject the null assumption of a common time-behavior before and after the announcement for information diffusion. In figure 2, the solid line denotes our estimate whereas the shaded area refers to our confidence intervals. Importantly, both information diffusion and attention to the news increase significantly in the hour after announcements.

Since we focus solely on announcements related to medical bulletins and policy measures to fight the epidemic, our results refer to both sources and topics distinct from those studied in the previous papers about economic announcements. Our results confirm that medical announcements gather special attention and hence it is important to understand whether they have a significant

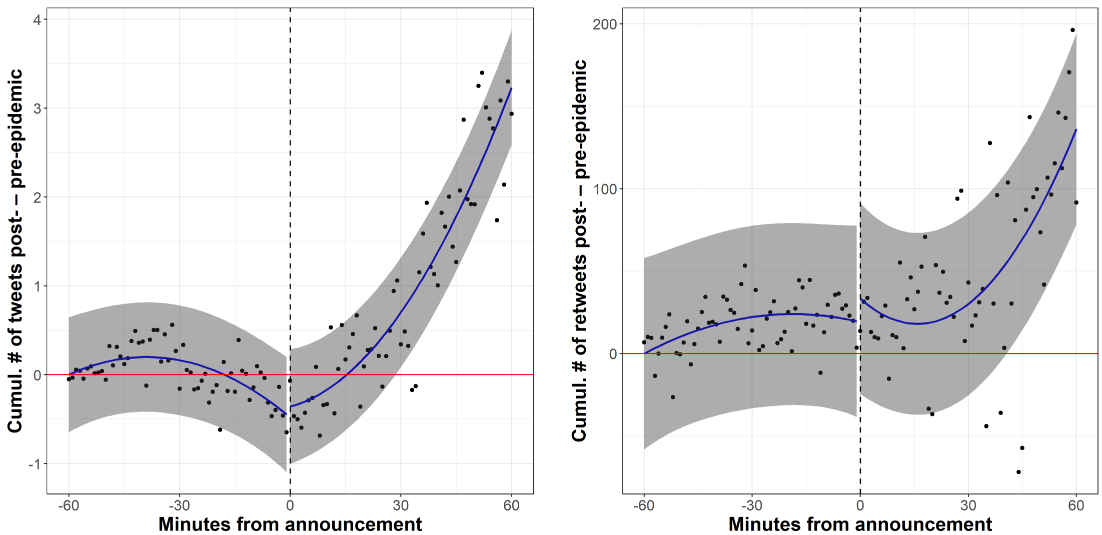


Fig. 2. Information Diffusion and Attention around Announcements

Notes: The left (right) panel of this figure shows the average per-minute and per-country number of tweets (retweets) around announcement times in excess of the same average in the pre-epidemic period. In each country, the epidemic period starts when there are more than 100 cases of COVID19. Solid line and shaded areas are based on the estimation of equation (1). The sample starts on October 1st 2019 and ends on the date of this draft.

impact on financial markets.

Financial data sources. All data are from Thomson Reuters and Bloomberg. Equity, bond and currency data are obtained at the minute frequency and then aggregated at lower frequencies when necessary. For each country, we collect data on its major equity index and 10-year maturity treasury bond index. We measure the risk-free rate by focusing on the yield of 3-month government bills. Due to data availability CDS data are collected at the daily frequency. All details about our data can be found in table A.3 (see appendix).

Equity markets. In figure 3, we show the average cumulative returns obtained from buying country-specific equities 60 minutes before a country-specific announcement and holding them for 120 minutes. Our results are averaged across both countries and announcements. Countries are divided in two groups, advanced and emerging economies, according to the IMF classification.[5]

The top panels show what happens when we consider all countries and all announcements. Namely, in AEs (EEs) equity values tend to slightly decline (stay flat) before the announcement and then appreciate substantially upon the announcement. This appreciation is persistent, as it remains almost constant during the next hour in AEs and it gets amplified in EEs. This observation suggests that the release of covid-related news helps equities. Since we are considering both announcements conveying positive news and announcements conveying negative news, we think of this jump in equity valuation as a measure of the value of the pure release of information about epidemic risk.

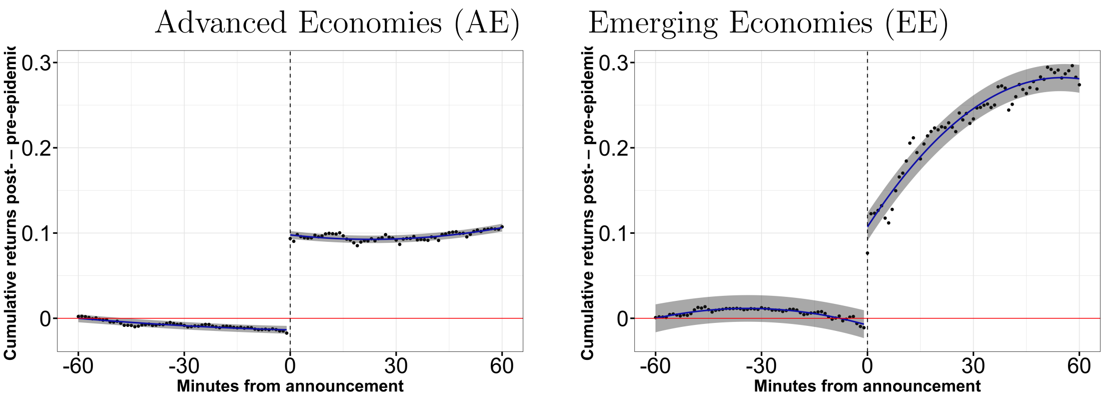
More specifically, we note that this figure shows a time varying behavior of returns that is quite different from that documented in Lucca and Moench (2015). Lucca and Moench (2015) show a slow and persistent accumulation of positive returns before monetary policy announcements. In our case, instead, the increase in the cumulative returns at the announcement is consistent with the Ai and Bansal (2018) model. When the representative investor cares about the timing of a resolution of uncertainty, prices jump upward when uncertainty is resolved along the information cycle, and then they eventually start to decline.

In figure 3(b), left panel, we show that the same phenomenon is present to a similar extent when we focus on the subset of announcements associated to bad news within the group of AEs.[6] We measure bad news as an unexpected increase in the growth rate of contagion cases on the day of the announcement. We explain in detail our construction of the news in the next section when we price them using the cross section of equity and bond returns.

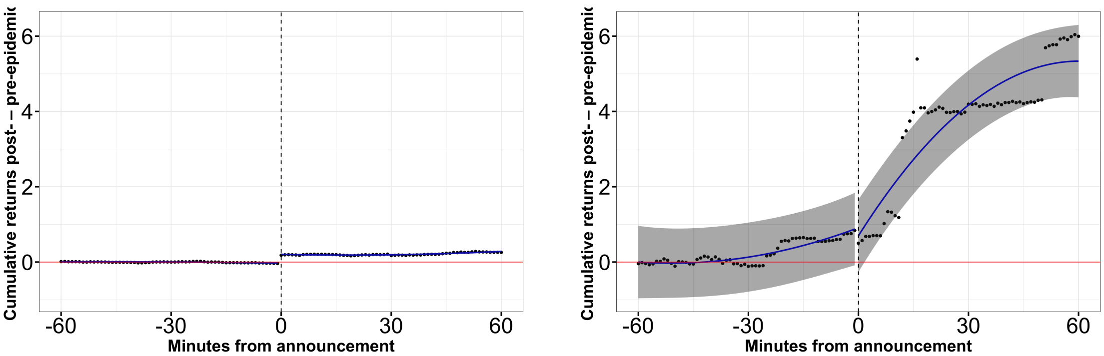
Turning our attention to EEs, we note that there still exists a positive jump in equity valuations, but it happens with about a 15-minute delay with respect to our announcement time stamps. Given our quadratic specification, this phenomenon is captured through a significant increase in the slope

[5]If a country-specific announcement happens when the exchange of the country is closed, we consider the 60 minutes prior to the closing time of the previous day and the first 60 minutes after the opening of the exchange in the next day.

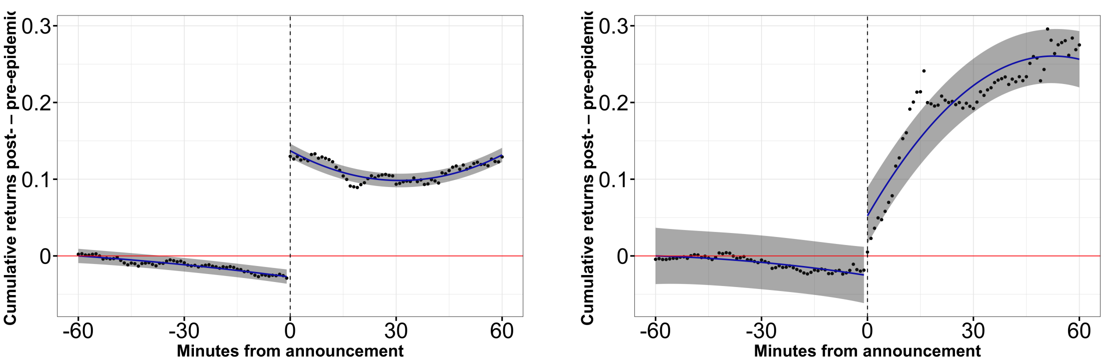
[6]Note that the scale for this panel is one order of magnitude grater than that in figure 3(a). Hence the announcement jump has the same magnitude as in panel a even though it looks smaller.



(a) All Countries



(b) Only Bad News



(c) High-COVID Countries

Fig. 3. Equity Returns around Announcements

Notes: In each panel, dots denote the difference across subsamples of the cross-country-cross-announcement average cumulative returns obtained from buying equities 60 minutes before an announcement and holding them for 120 minutes. Panel a (c) comprises announcements from all countries (top-50% countries in terms of contagion cases) in each group. Panel b excludes announcements conveying good news. Returns are in log units and multiplied by 100. Solid line and shaded areas are based on the estimation of equation (1). Our sample starts on October 1st 2019 and ends on the date of this draft.

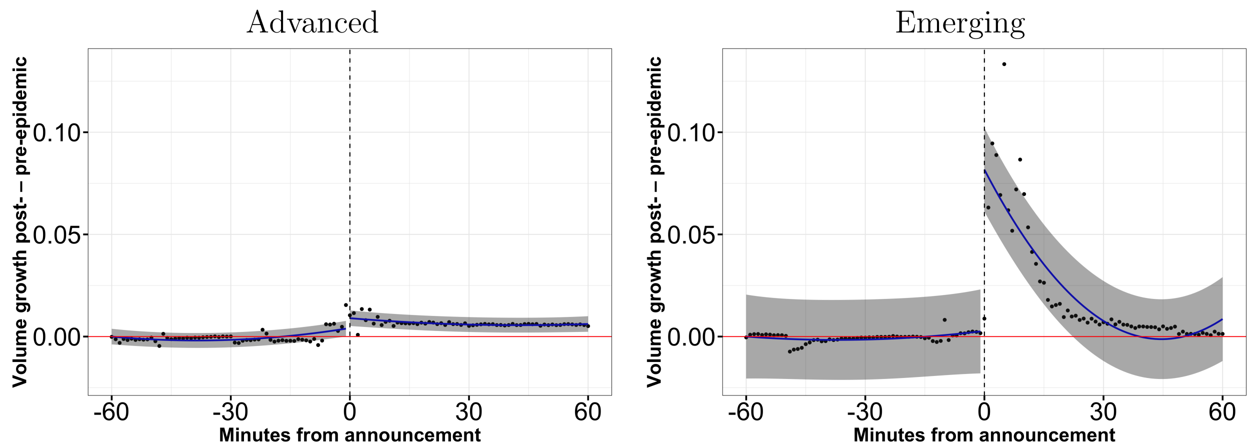


Fig. 4. Equity Volume around Announcements

Notes: The left (right) panel shows the average equity log-volume growth for all (above median of contagion cases) countries around announcement times. We depict the difference across pre- and post-epidemic samples. In each country, the epidemic period starts when there are more than 100 cases of COVID19. Solid line and shaded areas are based on the estimation of equation (1). Our sample starts on October 1st 2019 and ends on the date of this draft.

of our cumulative returns time series. We also point out that in this case the jump is one order of magnitude greater than under the case in which we consider all announcements, implying that in these countries the value of resolution of uncertainty may be extremely high even when we condition on bad news.

In figure 3(c), we consider all of our announcements but we limit our attention to countries that are above median in terms of total contagion cases. The scale in these panels is identical to that used in figure 3(a). Not surprisingly, the smaller sample that we use produces estimates surrounded by higher estimation uncertainty. Taking this into account, the value of the information disclosed during these announcements is higher among high-COVID AEs and remains almost unchanged among high-COVID EEs. More broadly, when we look at the entire cross section of our 21 countries, low-COVID countries appear to be less sensitive to contagion-risk news. This is consistent with the results of the no-arbitrage factor model that we estimate in the second part of our study.

The equity returns patterns that we document may also be consistent with models featuring behavioral attributes and micro-frictions. In order to provide more data to distinguish across theories, we also look at equity volume. In figure 4, we directly depict the difference in volume log-

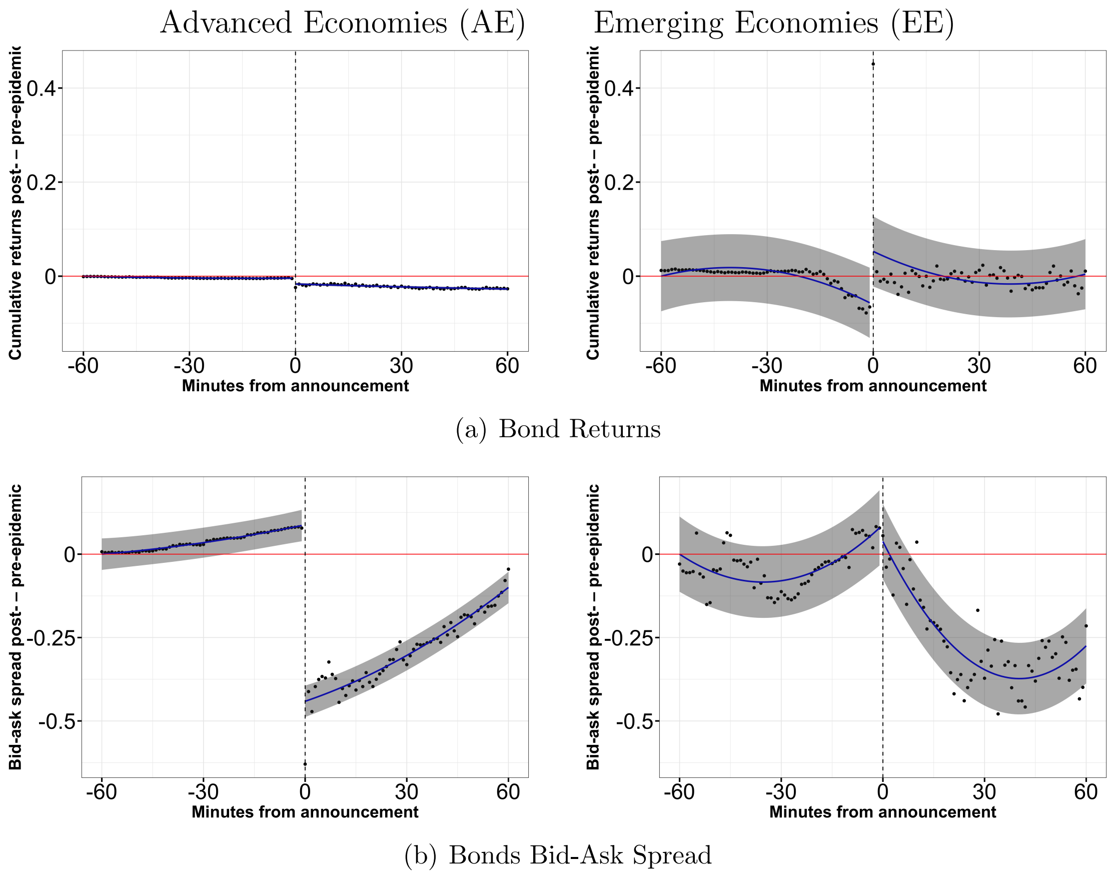


Fig. 5. Sovereign Bonds around Announcements

Notes: In the top panels, dots denote the difference across subsamples of the cross-country-cross-announcement average cumulative returns obtained from buying 10-year sovereign bonds 60 minutes before an announcement and holding them for 120 minutes. In the bottom panels, dots refer to the difference across subsamples of the cross-country-cross-announcement average of the bid-ask spread of the bonds. Returns are in log units. All series are multiplied by 100. Solid line and shaded areas are based on the estimation of equation (1). Our sample starts on October 1st 2019 and ends on the date of this draft.

growth across normal and epidemic subsamples. We find that both in AEs and in EEs trade volume features no change before the announcements. Consistent with previous studies (see, among others, Han (2020)), trade volume increases right after the announcement. This upward adjustment is more pronounced in EEs. In the next part of this study, we focus on sovereign bonds and document that liquidity seems to increase in the bond markets as well.

Bond markets. Figure 5(a) shows our results for bonds returns. The construction of the depicted data is identical to that used for equities. We note that the dynamics in the bond markets are less severe than those observed from equities. In a $\pm$60-minute window around the announcement, there is no significant adjustment in bonds returns for AEs. This observation is important as, by no-arbitrage, it suggests that cash-flow uncertainty is an important determinant of the market fluctuations observed during the COVID19 crisis. This high-frequency result is consistent with the results documented by Gormsen and Koijen (2020) looking at dividend futures.

Focusing on EEs, however, we note that sovereign bonds loose value ahead of announcements and then appreciate at the time of announcement like equities. Over our $\pm 60-$minute window, however, the cumulative return is nearly zero both across AEs and EEs, suggesting that bonds are an important hedge against contagion risk announcements.

In order to further investigate the role of sovereign bonds, we also look at the behavior of their bid-ask spread. Absent high-frequency data on bonds trading volume, we think of this spread as a measure of liquidity in the market. We note an immediate decline in the bid/ask spread in AEs and a delayed one in EEs. This observation, paired with the decline in equity volume depicted in figure 4, suggests that investors may tilt their trade toward bonds right after announcements. In AE countries, we should not be surprised that such a reallocation of investment flows comes with almost no adjustment in bond prices since it may be the result of their monetary policy.

An alternative explanation for this muted response is that bond markets are subject to two offsetting forces. Specifically, flight to safety may promote bond appreciation but, simultaneously, sovereign default risk may increase and push bond prices downward. In order to study the plausibility of this hypothesis, we collect daily country-level data on CDS spreads and link their daily variation to daily news on contagion cases. We explain in detail how we measure news in the next section. Given that different countries entered this crisis with different levels of fiscal capacity, exploring country-level heterogeneity is important. Fore this reason, in our empirical analysis we include both country-level fixed effects and week-level time fix effects.

Table 2. CDS Spreads and Contagion News

	A.E.		E.E.	
Contagion cases - news	6.269***	8.066**	28.889***	28.970***
	(2.066)	(3.988)	(8.614)	(8.891)
Adj. R2	0.01%	4.75%	0.17%	14.30%
Adj. R2 w/o	0.01%	4.75%	0.17%	14.30%
Country FE	Yes	Yes	Yes	Yes
Week FE	No	Yes	No	Yes

Notes: this table reports the results of the following regression:

$$\Delta S_t^i = d_0^i + d_t^i \cdot D_t^{Week} + \beta^g \cdot news_{t-1} + \epsilon_t^i, \quad \forall i \in g$$

where ΔS_t^i refers to the daily change of the CDS spread in country i; g refers to either the group of Advanced Economies (AEs) or that of Emerging Economies (EEs); d_0^i is a country-level fixed effect and D_t^{Week} is a weekly time fixed effect. 'Contagion cases - news' refers to the innovation in the growth of the global number of contagion cases as measured in section 3. 'Adj. R2 w/o' refers to the adjusted R squred from the same regression in which we omit the contagion news. Standard Errors are clustered at the country-level. Our sample starts on October 1st 2019 and ends on the date of this draft.

In table 2, we show that that adverse contagion news tend to increase CDS spreads in a statistically significant way. This effect is three times stronger in EEs. Simultaneously, we document that these news produce a very modest increase in the adjusted R-squared of our regression, implying that for AEs, default concerns have been a second-order issue.

The role of domestic announcements. Recall that our cross section comprises 21 countries. We can think about the previous results about equity (bond) returns as the equal-weighted cumulative returns that an investor could obtain by trading ahead of each announcement across 21 sources of announcements (one per country) and in 21 equity (bond) markets, for a total of 21^2 possible trade combinations.

In order to disentangle the effects of local announcements on local markets, we also consider the average cumulative return of an investor that trades only in the domestic market ahead of domestic announcements. In figure 6, we focus on the average cumulative returns across 21 trade strategies that involve neither foreign news nor foreign assets. Our data confirms that bonds have a muted

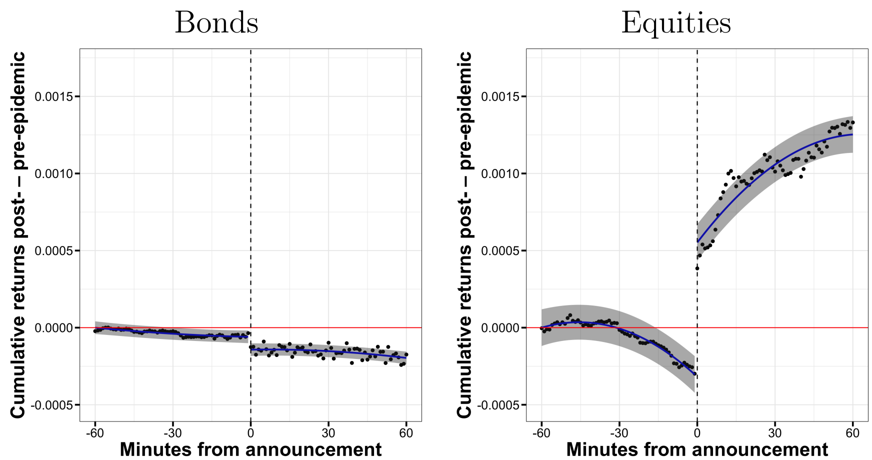


Fig. 6. Local Returns around Domestic Announcements

Notes: In each panel, dots denote the difference across subsamples of the cross-country average cumulative returns obtained from buying domestic equities 60 minutes before a domestic announcement and holding for 120 minutes. Returns are in log units. Solid line and shaded areas are based on the estimation of equation (1). Our sample starts on October 1st 2019 and ends on the date of this draft.

response to announcements. Equities, in contrast, tend to depreciate ahead of the announcement and then suddenly appreciate afterward. This pattern resembles that derived by Ai and Bansal (2018) in a model in which the timing of information matters.

3 Contagion News

In this section, we attempt to price news about pandemic risk. We do it using two fundamental measures, namely, unexpected change in number of contagion cases and unexpected change in the tone of the news about contagion. The first measure is based on an objective count of COVID19 positive cases. Yet, across different months or contagion waves, the same variation in the number of cases may be associated to different assessments of risk. For this reason, we find it important to study also a media-based measure of news tone.

Our analysis confirms that global epidemic news have a significant market price of risk. In April 2020, at the peak of the first COVID contagion wave in AE, daily equity risk premia may have increased by 28% in AEs and by 13% in EEs compared to the median risk premia in our sample.[7]

3.1 Data Collection

Twitter-based news. In the spirit of Baker et al. (2016), we identify major newspapers for a large cross section of countries (see table A.1 in the appendix). In contrast to Baker et al. (2016), we do not analyze articles, rather we track news published on Twitter in real time, so that we can produce high frequency data when needed. More specifically, we track the news related to the COVID19 pandemic posted by major newspapers on Twitter. We do so by searching for key words such as 'coronavirus' and 'covid19'. For each newspaper, we identify the location of its headquarter so that we can identify its specific time-zone.

In table 3, we report a summary of our social media–based dataset. It is very comprehensive and it features several dimensions that enable us to study both information production and diffusion. Specifically, our ability to track retweets and likes gives us a high-frequency measure of attention. Google searches are often used to measure attention (Da et al. 2011; Ramelli and Wagner 2020), but to the best of our knowledge they are not provided minute-by-minute and they do not account for the timing of initial production of the news, an aspect that is very important when analyzing capital market reactions.

The time series behavior of our news indicator is depicted in figure 7. For each country, we also depict the beginning of the epidemic period which we identify on the day in which the number of confirmed cases of COVID19 becomes greater than 100. We note several interesting patterns. First of all, there is significant heterogeneity across countries in the timing of the information diffusion. Across several countries, information diffusion becomes more intense after the beginning of the local epidemic period. We note that both the diffusion of news, that is, number of tweets, and the attention to the news, that is, number of retweets, increase rapidly after the beginning of the local

[7] These numbers are annualized according to the number of annual trading days and are net of the median risk premium in our full sample.

Table 3. Newspapers Dataset

Country	No. News Providers	Tweets	Retweets	Likes	Topics			
					Mortality	Quarant.	Med. Supply	Vaccines
Argentina	4	72,122	1,154,418	3,005,813	13%	10%	15%	62%
Australia	4	15,149	132,995	307,911	17%	45%	12%	27%
Brazil	4	30,482	1,303,013	8,441,881	46%	8%	16%	30%
Canada	5	42,194	399,062	749,014	33%	11%	17%	40%
Chile	4	30,822	384,121	582,193	54%	6%	11%	29%
China	3	30,756	922,958	2,532,730	41%	15%	19%	25%
Colombia	4	29,862	454,434	1,383,847	18%	13%	25%	44%
France	4	43,032	1,322,400	2,198,635	25%	28%	27%	20%
Germany	4	11,074	138,910	295,222	17%	26%	21%	36%
Hong Kong	3	19,256	409,874	588,364	18%	32%	22%	29%
India	4	97,408	905,550	5,385,515	33%	23%	17%	27%
Italy	3	31,854	259,334	697,363	11%	32%	29%	27%
Japan	4	16,935	146,280	252,278	18%	13%	31%	38%
Korea	4	11,671	75,204	130,382	41%	11%	29%	20%
Mexico	4	74,038	1,551,902	4,015,242	15%	12%	25%	49%
New Zealand	4	22,508	168,236	401,181	21%	40%	15%	24%
Spain	4	36,888	2,561,751	4,552,308	31%	21%	14%	34%
Switzerland	4	7,773	35,965	44,589	23%	20%	24%	33%
UK	4	24,096	1,121,290	2,226,512	27%	31%	15%	27%
USA	11	105,285	6,876,316	16,029,160	30%	8%	23%	40%
Total	85	753,205	20,324,013	53,820,140	27%	20%	20%	33%

Notes: This table shows summary statistics of COVID19-related news data that we collect for a large cross section of countries. Our real-time data range from January 1st 2020 to the date of this manuscript. For each country, we report number of news providers and number of tweets collected. We also report the total number of retweets and likes as measures of attention. The last four columns report the share of tweets mentioning number of deaths, quarantine measures, medical supply, and vaccines, respectively.

epidemic period.

Figure 8 shows both diffusion and attention to the news at the global level, that is, when we aggregate all of our tweets and retweets across countries. The right panel of this figure provides a breakdown of the most prominent topics addressed in the COVID19 tweets, namely, vaccines, death risk, quarantine measures, and availability of medical supply. The attention to all of them increased substantially, with vaccines becoming prominent in the fall 2020.

Figure 9 shows the intraday pattern of the diffusion of COVID19 news for each country. This figure is not based on universal time, rather it accounts for country-specific time. In each country, we consider two country-specific subsamples, that is, the pre-epidemic and epidemic period. There are two main takeaways from this picture: (i) the diffusion of COVID19-related news increases

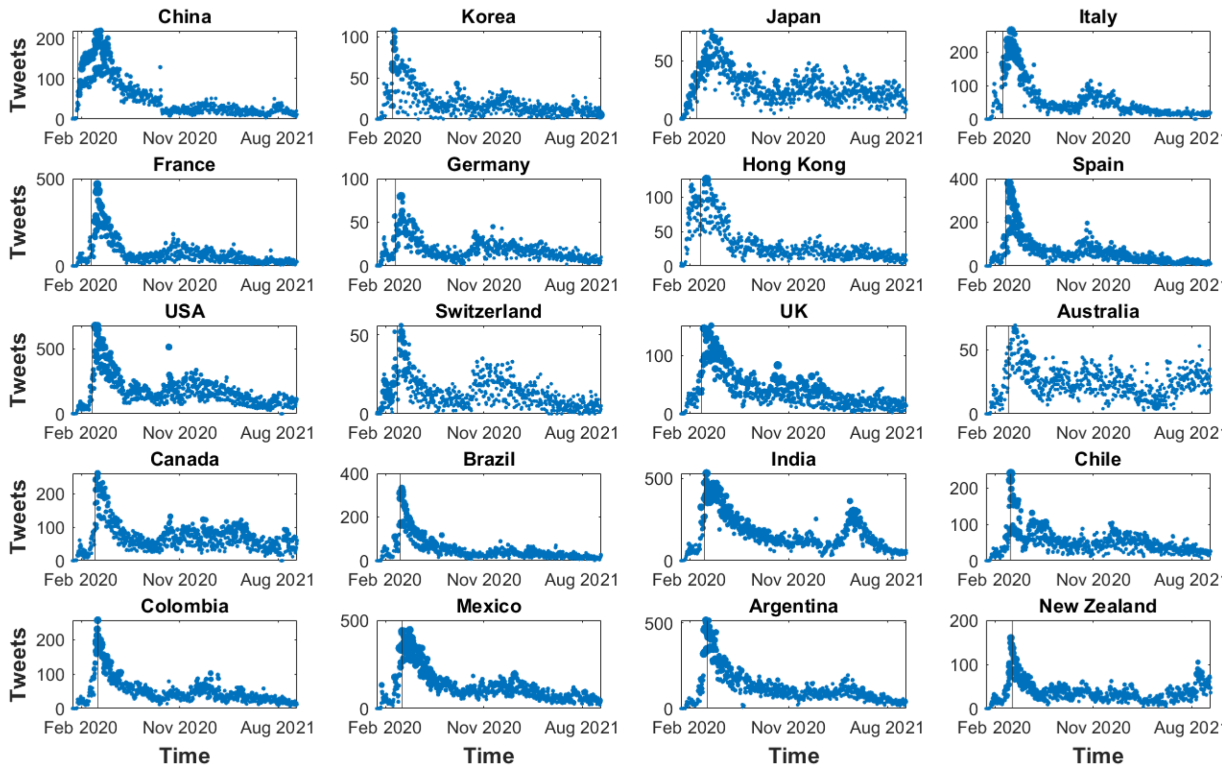
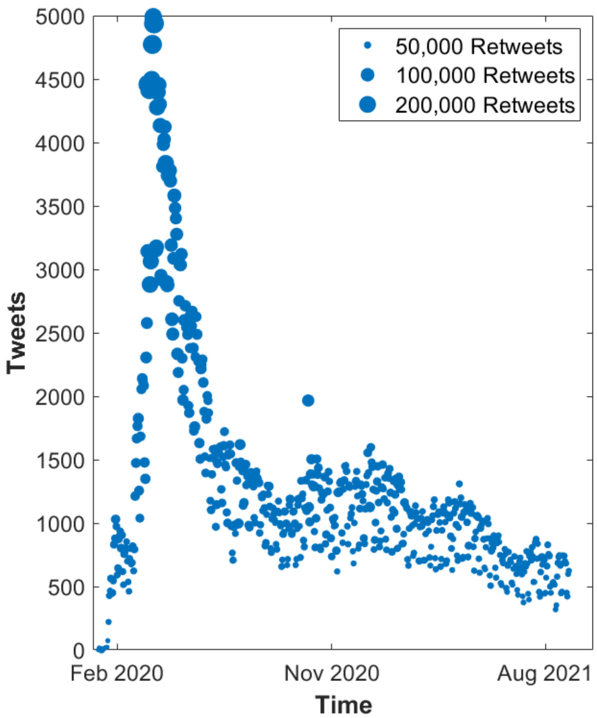


Fig. 7. Information Diffusion and Attention across Countries

Notes: This figure shows the daily number of tweets posted in each country by major newspapers. The vertical axis shows the daily number of tweets. The size of each data point represents the number of retweets scaled by the maximum daily number of retweets for each country. The sample starts on January 8th 2020 and ends on the date of this draft. The vertical line depicts the date that each country had more than 100 confirmed cases of COVID19. More details on the data collection are reported in the Appendix.

significantly with local epidemic conditions; and (ii) a significant share of the diffusion takes place while the local capital markets are open. Hence monitoring media activity can be a very useful tool to track in real-time the information set of financial market participants.

Tweet Tone. Since we use Twitter activity to form a high-frequency risk factor, we need to identify the tone of the tweets, that is, we need to know whether they relate to either good or bad news. Given (i) the high volume of tweets that we collect, and (ii) the fact that our tweets are written in different languages, we use Polyglot (available at `https://pypi.org/project/polyglot/`), i.e., a natural language pipeline that supports multilingual applications with polarity lexicons for 136



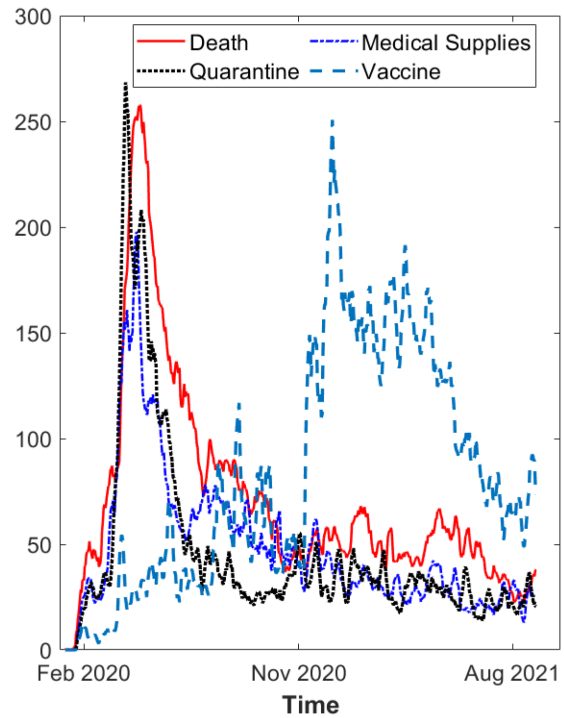


Fig. 8. Global Information Diffusion

Notes: The left panel of this figure shows the daily total number of tweets posted across countries by major newspapers. The vertical axis shows the daily number of tweets. The size of each data point represents the number of retweets scaled by the maximum daily number of retweets. The right panel shows the daily number of tweets related to death-risk, (scarcity of) medical supplies, quarantine, and vaccines. The tweets were identified using a multilingual bag-of-words approach. The sample starts on January 8th 2020 and ends on the date of this draft. More details on the data collection are reported in the Appendix.

languages. This computer-based mapping algorithm reads our text and classifies the words into three degrees of polarity: +1 for positive words, -1 for negatives words and 0 for neutral words. We provide two examples in table A.2 (see our appendix).

Our measure of the tone of the tweets is based on the count of positive words minus the count of negative words, divided by the sum of positive and negative word counts (Twedt and Rees, 2012). We compute this measure at the country level at both the hourly and the daily frequency. We then aggregate this measure across countries in order to obtain a global measure.

We depict our global tone factor in figure 10, left panel. Its time-pattern is consistent with the observed contagion dynamics. Specifically, the tone became very negative by the end of January as

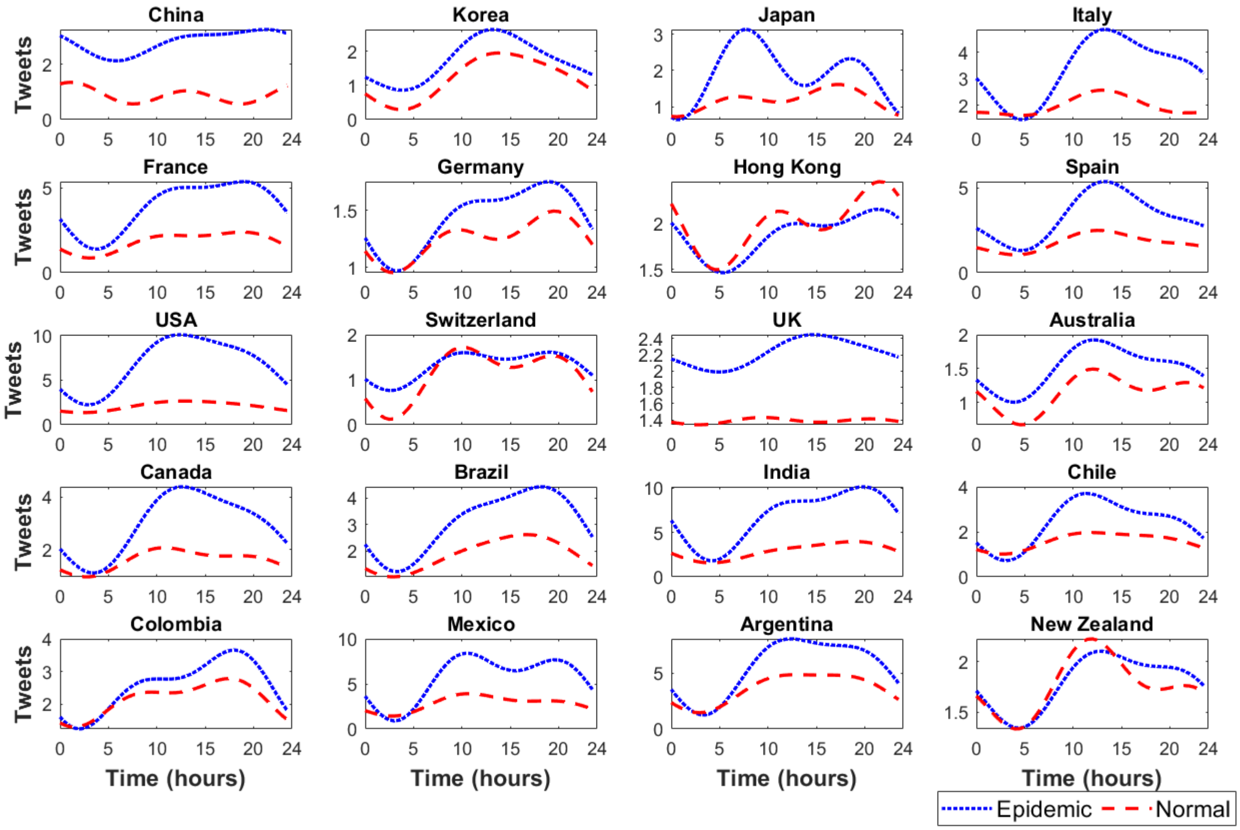


Fig. 9. Intraday Information Diffusion

Notes: This figure shows the intra-day trend of the number of tweets posted every 30 minutes across several countries in our dataset. The dotted line represents the intra-day trend in the epidemic period, identified when a country has more than 100 cases of COVID19. The dashed line represents the intra-day trend in the pre-epidemic period. The sample starts on January 8th 2020 and ends on the date of this draft. Time refers to local time zone of each newspaper. More details on the data collection are reported in the Appendix.

the conditions in China started to precipitate. It improved in early February, when there was still no sign of massive contagion in Europe, and it declined again when the epidemic started in Italy. The slow improvement of the tone of our tweets observed after the beginning of March pairs well with the observed flattening of the contagion curves in many of the countries in our dataset. We find these results reassuring as they confirm that our text analysis algorithm tracks the contagion dynamics in a reliable manner.

For the sake of our asset pricing analysis, we focus on the innovations to the tone of our tweets.

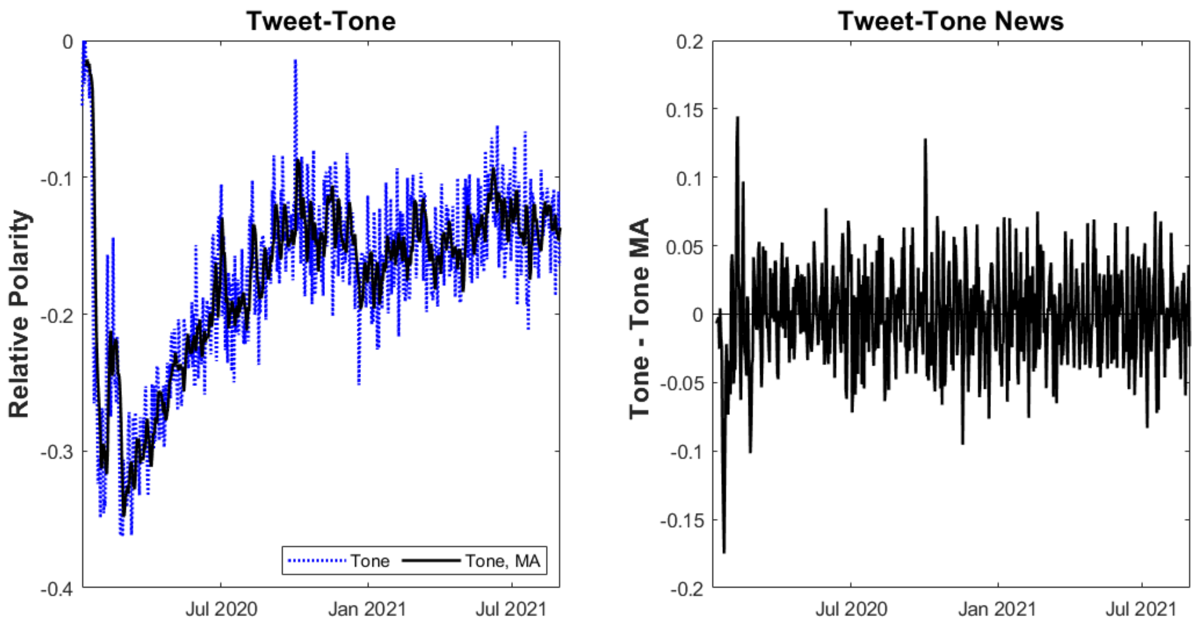


Fig. 10. Twitter-Based COVID19 Factor

Notes: This figure shows our daily global Twitter-based COVID19 factor. We use Polygot to measure the polarity of our tweets and compute the tone of each tweet according to Twedt and Rees (2012). We aggregate the tones at a daily frequency and across countries. MA refers to a backward looking 5-day moving average. The news at time t is computed as the difference between the tweets-tone at time t and their MA at time $t-1$. The sample starts in early January 2020 and ends on the date of this draft.

One simple way to extract these innovations is to consider the difference in the tone at day t and its 5-day backward looking moving average assessed at time $t-1$. We depict this time series in the right panel of figure 10 and note that it is nearly serially uncorrelated.

Contagion and financial data. Contagion data are from official medical bulletins. Our primary source is CSSE at Johns Hopkins University.[8] News to the contagion factor are obtained by computing the difference between the daily growth rate of contagion cases at time t and its backward-looking time $t-1$ moving average computed over the previous 5 days. We choose a 5-day window because it matches the number of days of a typical trading week.

Since our contagion-based factor spans a 7-day week, we assign to Friday the average growth

[8] https://github.com/CSSEGISandData/COVID19/tree/master/csse_covid_19_data/csse_covid_19_time_series

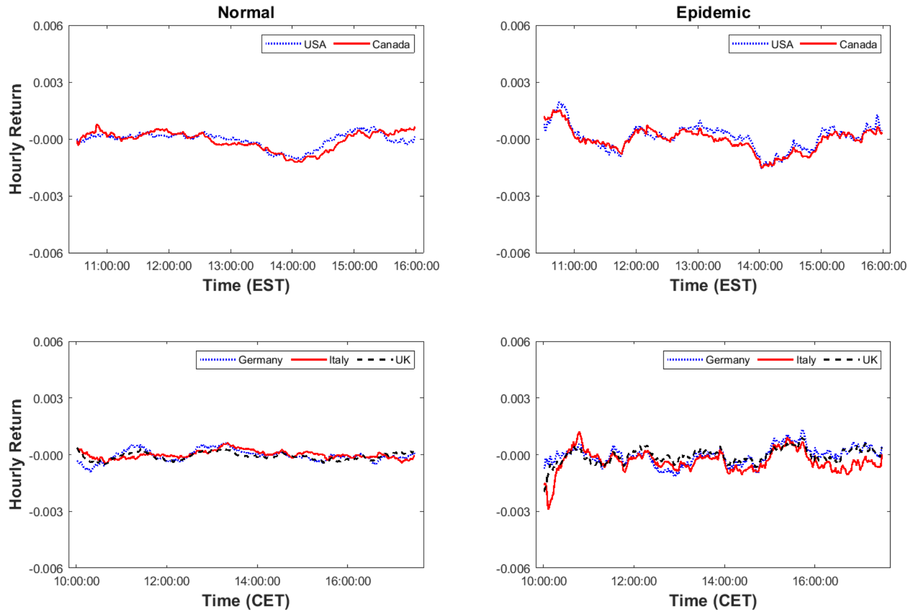
rate of global contagion cases that occurred on Friday, Saturday, and Sunday.[9] Our financial data sources are detailed in table A.3 (see appendix).

In order to show the relevance of local epidemic conditions, in figure 11 we show the intra-day behavior of returns pre- and post-epidemic for equities, bonds, and currencies. We focus on two groups of countries with similar stock exchange timing, namely US and Canada (EST timezone), and Italy, UK, and Germany (CET timezone). The countries in the second group are interesting because they have experienced very different exposures to COVID19. Italy has been affected first and in an intensive way. Germany has been able to mitigate the contagion during the first contagion wave and has seen a pick up in contagion numbers as soon as it lessened the lockdown measures. The UK has changed its strategic response to the crisis in the middle of the epidemic period.

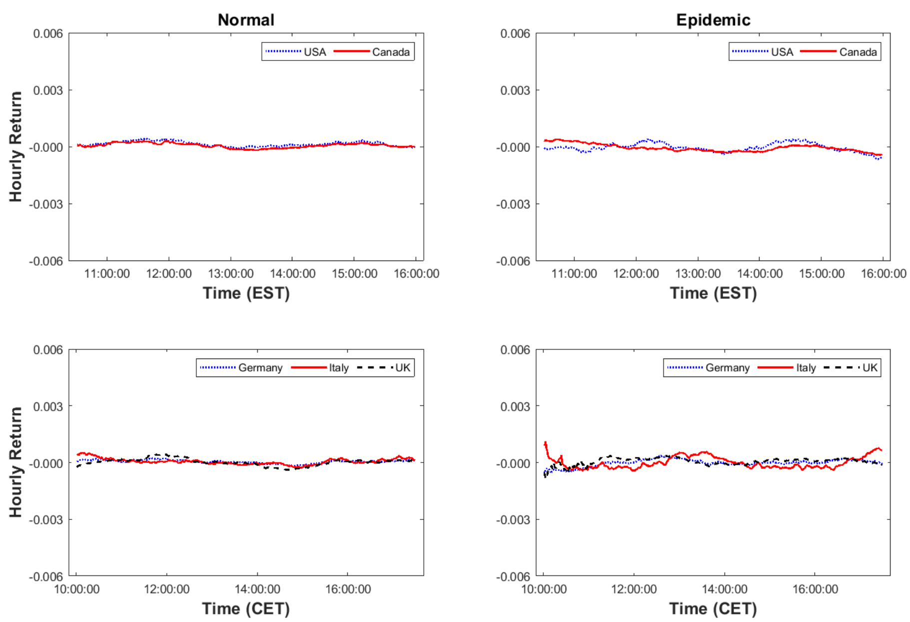
We note that equity returns have been much more volatile in the epidemic period. Most importunately, the intra-day patterns have become much more correlated once all countries have gone into an epidemic state. This result suggests that we can think of the epidemic as a slowly diffusing common factor. Our empirical asset pricing analysis is based on this observation.

When we turn our attention to bonds in the epidemic period, we see more volatile patterns than in the pre-epidemic period. In contrast to equities, we see no substantial change in their commonalities across countries. Currencies, instead, tend to be more volatile and more correlated in epidemic subsamples, similarly to equities. We see this as consistent with COVID19 being a global risk factor that affects countries at different times and with different intensities. Our empirical asset pricing analysis takes into consideration the hypothesis that our countries may feature heterogeneous exposure to global contagion risk.

[9] For the Easter Holiday, we assign to Thr the average daily growth rate of global cases from Thr to the following Mon.



Equities



Bonds

Fig. 11. Intra-day Returns Behavior and Epidemic Conditions

Notes: For each asset class, we depict per- and post-pandemic intra-day return patterns. Data are averaged across days. In each country, the epidemic period starts when there are more than 100 cases of COVID19. The sample starts in October 2019 and it ends October 2020. Bond and stock hourly returns start one hour after the opening of the markets. All returns are in raw units.

Table 4. **Summary Statistics for Portfolios**

	Low	Medium	High	HML$_{COVID19}$
Panel A: Advanced economies				
Mean	0.027	0.054	0.023	-0.004
	(0.072)	(0.072)	(0.089)	(0.039)
StDev	1.241	1.43	1.533	1.115
Skewness	-1.234	-0.778	-1.642	-0.099
First Quartile	-0.463	-0.45	-0.525	-0.566
Avg. N. Countries	5.008	4.002	4.99	-
Turnover (%)	0.5	1.2	0.6	-
Panel B: Emerging economies				
Mean	0.008	0.021	0.112*	0.104
	(0.107)	(0.112)	(0.057)	(0.073)
StDev	1.829	1.938	1.843	1.668
Skewness	-2.056	-1.357	-0.812	0.383
First Quartile	-0.697	-0.887	-0.783	-0.956
Avg. N. Countries	3.003	1.997	2	-
Turnover (%)	0.5	1	0.6	-

Notes: This table shows summary statistics for the equity excess returns of portfolios formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation. Hourly excess returns are in log units and multiplied by 100. Portfolios are obtained from equity indexes. Our real-time data range from February 2020 to the date of this manuscript. Turnover measures the number of countries entering or exiting a portfolio relative to the total number of countries in a specific portfolio $\times$ number of days in our sample. Numbers in parenthesis are HAC-adjusted standard errors.

3.2 The Market Price of Viral Contagion News

Daily news. Every day, we group countries into three portfolios according to their relative number of COVID19 cases measured the previous day. We do this separately for AEs and EEs. The H (L) portfolio comprises the top (bottom) countries in terms of COVID19 cases. We also consider an investment strategy long in the H portfolio and short in the L portfolio. We refer to the returns of this portfolio as HML-COVID19.

We report common summary statistics for these portfolios in table 4. The turnover in each portfolio is moderate. The in-sample average of the returns in all portfolios is not different from zero, which is not surprising given our short sample which comprises both the first contagion wave and its temporary disappearing. All portfolio returns have substantial volatility and negative skewness. Focusing on the first quartile of the distribution of returns, we see that the portfolio comprising the more exposed countries tends to have more severe negative downside risk. This is an aspect that we capture in our conditional no-arbitrage model.

Given these preliminary observations, we consider the following conditional asset pricing model,

$$r^{ex}_{f,t+1} = \overline{r}^{ex}_{f,t} + \beta_{f,t} \cdot news^{glob}_{t+1}, \quad f \in \{H, M, L\}, \tag{2}$$

$$\beta_{f,t} = \beta_0 + \beta_{f,1} X_{f,t}, \tag{3}$$

$$\frac{\partial \overline{r}^{ex}_{f,t}}{\partial X_{f,t}} = \lambda \beta_{f,1}, \tag{4}$$

where X_t is the share of contagion cases associated to portfolio f at time t, and λ is the market price of risk (MPR) of the global news factor $news^{glob}_{t+1}$.

This model can potentially capture many of the features of returns seen so far. First, it captures predictability through contagion-based time-varying betas. Second, it has the potential to capture higher negative skewness for countries that go through more severe contagion paths. Consider the case of portfolio H comprising countries receiving a sequence of relatively more severe adverse contagion news. This portfolio will have severe exposure to adverse news as the relative contagion share of the portfolio grows. When the relative contagion share starts to flatten out and decline, the sensitivity of this portfolio to good news is reduced ($|\beta_{H,t}|$ shrinks). This means that returns become less sensitive to positive news and hence the right tail of the returns distribution is shortened.

Third, consistent with our previous descriptive returns, it accounts for heterogeneous exposure to global contagion news. Last but not least, it enables us to identify the market price of risk of this global contagion component, λ. By no-arbitrage, the extent of time-series predictability of our excess returns must equal $\lambda\beta_{f,1}$, and $\beta_{f,1}$ can be easily estimated in the time-series by considering the multiplicative factor $X_{f,t} \cdot news^{glob}_{t+1}$.

We report our main results obtained from daily data in table 5. Panel A is based on unexpected changes in the growth of global contagion cases. Panel B, instead, is based on unexpected changes in the global tone of tweets. Note that the set of countries that we consider provide daily updates about contagion cases at the end of the day. In order to properly represent the information set of investors, in our asset pricing model we lag the news by one day, i.e., we assume that day-t returns respond to news released in the evening of day $t-1$.

We estimate our asset pricing model through GMM and notice that all portfolios have an untabulated significant exposure to our contagion-based news, $\beta_{f,t}$.[10] In our sample, the portfolio of countries with the highest share of COVID19 cases tends to be more exposed to contagion news. This sign is consistent with our expectations since positive (negative) news about global contagion growth (tone of tweets) refers to an adverse shock to equity returns. Most importantly, the implied daily market price of risk is negative (positive) and significant with respect to contagion (tone of tweets) news. This means that the relative share of contagion cases forecasts an increase in expected future returns across all portfolios ($\lambda\beta_{f,1} > 0$). Equivalently, the share of contagion cases is a relevant positive predictor of future cost of capital.

Our results hold regardless of whether we run our model using local-currency returns or returns in USD. Furthermore, our results remain significant when we estimate a two-factor version of our model which controls by global market risk as measured by the MSCI Global Index.[11] This result holds both when we use only equities as test assets and when we increase our cross section by introducing bonds. Looking at the output of our specifications and accounting for estimation uncertainty, we conclude that 0.3% is a reasonable lower bound on the daily market price of risk of daily contagion news. We consider this estimate as very significant, consistent with the great contraction experienced in equity markets during the first waive of the epidemic period.

Simultaneously, we note that this value is very plausible once we account for two observations.

[10]The share of contagion cases across our three portfolios have very different scales and variability. As a result, the coefficients $\beta_{f,1}$ are not revealing of the sorting of $\beta_{f,t}$ across portfolios. For this reason, we report only estimated MPRs.

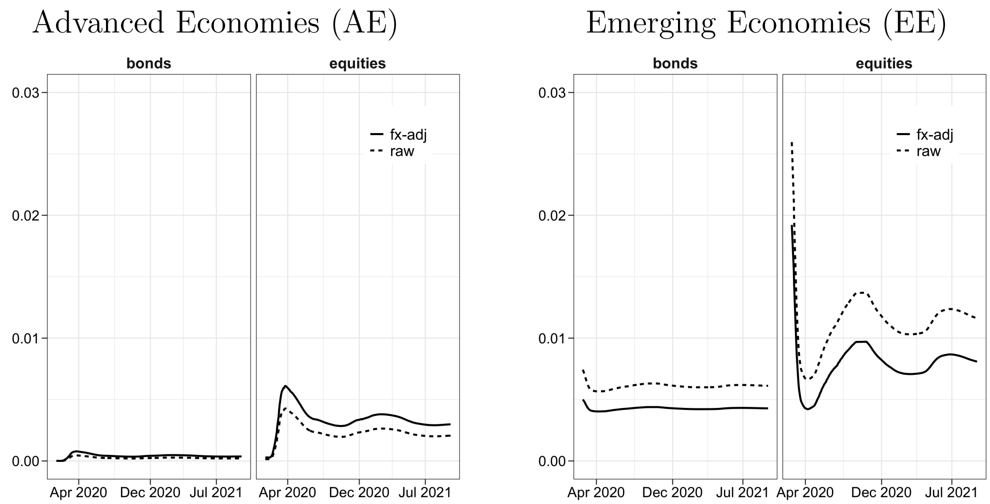
[11]Throughout our study, when we consider the MSCI index to control for the market we use returns in USD.

Table 5. Summary of MPR estimation

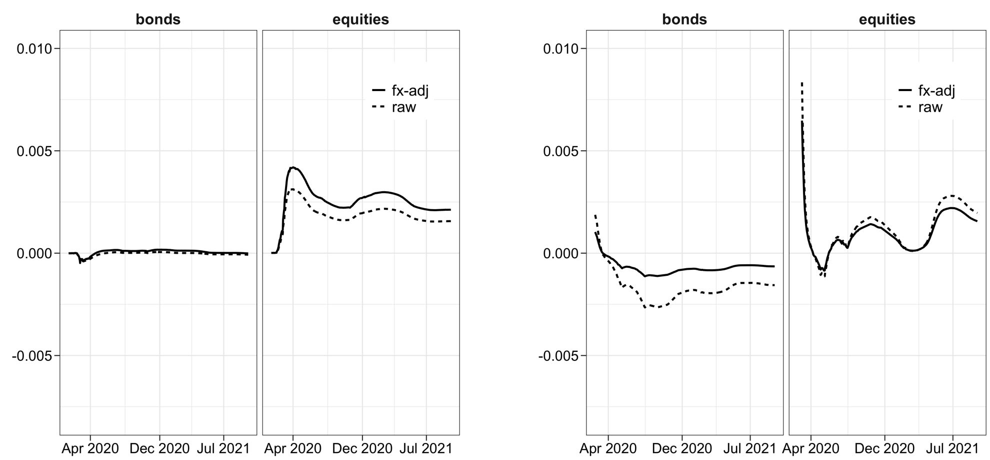
	Equity		Bonds & Equity	
	A.E.	E.E.	A.E.	E.E.
Panel A: News about Covid cases				
Local units				
coef	-0.012^{***}	-0.005^{**}	-0.003^{***}	-0.005^{***}
se	(0.005)	(0.002)	(0.001)	(0.001)
USD units				
coef	-0.013^{**}	-0.005^{***}	-0.006^{***}	-0.004^{***}
se	(0.006)	(0.002)	(0.001)	(0.001)
Controlling for MKT				
coef	-0.002	-0.003^{***}	-0.003^{***}	-0.007^{***}
se	(0.002)	(0.001)	(0.001)	(0.001)
Panel B: News from Twitter				
Local units				
coef	0.027^{***}	0.018^{***}	0.026^{***}	0.006^{***}
se	(0.008)	(0.004)	(0.004)	(0.001)
USD units				
coef	0.026^{***}	0.011^{***}	0.015^{***}	0.007^{***}
se	(0.007)	(0.002)	(0.003)	(0.001)
Controlling for MKT				
coef	0.001	0.018^{***}	0.007^{***}	0.006^{***}
se	(0.003)	(0.003)	(0.001)	(0.001)

Notes: This table shows the results of the conditional linear factor model described in equations (2)–(4). Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). In panel A (panel B), the COVID19 factor is measured as the news to global COVID cases growth (tone of COVID-related tweets). When we measure the COVID19 news as unexpected number of contagion cases (unexpected improvement in COVID19-related tweets), we expect a negative (positive) market price of risk (MPR). Both daily excess returns and market prices of risk are in log units. The last two columns are based on a broader cross section of test assets comprising both equity and bond portfolios. When we control for the market, returns are in USD, the market is measured by the MSCI Global Index and our factor model comprises a total of two factors. Our real-time data range from 2/1/2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

First, this is not the MPR of a financial factor and the associated estimated beta are very small. Second, contagion risk follows waves with a relatively short half-life. Equivalently, the exposure of our assets to this risk are small and relatively quick in reverting to zero. This phenomenon is depicted in figure 12(a). Our results confirm that sovereign bonds issued by AEs are not sensitive



(a) Expected returns for H_{COVID} portfolios



(b) Expected HML_{COVID}

Fig. 12. Expected Risk Premia

Notes: The left (right) panels refer to portfolios of countries within the AE (EE) group. The top panels show the estimated risk premium on a portfolio of countries with a share of High-COVID19 cases on bond and equity portfolios. The bottom panels refer to the HML-COVID strategy. These results are based on the specifications reported in the last two columns of table 5. The solid line refers to exchange rate–adjusted returns, i.e., returns expressed in USD.

to contagion risk. Equities, instead, experienced a more pronounced increase in their required risk premium among High-COVID countries. In contrast, in EEs both bonds and equities feature a much more pronounced increase in their riskiness. Bonds' exposure, however, has been smaller than that of equities', confirming that also EEs' bonds are safer with respect to contagion risk.

In figure 12(b), we show the estimated risk premium on an HML-COVID19 strategy on either bond or equity portfolios across AEs and EEs. Focusing on this strategy helps us to highlight the role played by heterogeneous exposure to contagion risk. We document several novel empirical results. First of all, we note that the riskiness of bonds has increased less in High-COVID countries than in Low-COVID countries. Equivalently, in High-COVID countries, bonds are relatively safer assets. As a result, an HML-COVID strategy on bonds provides an insurance premium. In AEs, this premium is very moderate, consistent with our prior empirical evidence on the muted response of bonds around medical announcements time. In EEs, instead, the insurance premium is quantitatively relevant both in local units and in USD. Hence this HML strategy may be of interest to international investors seeking a strong hedge against contagion risk.

Second, we notice that the equity-based HML strategy in AEs features a required premium similar to that estimated for the High-COVID portfolio. Equivalently, Low-COVID countries have experienced nearly zero change in their risk premium. This result is important because it implies that containment policies that keep contagion cases relatively low may be very valuable both in terms of lives saved and in terms of preventing severe financial wealth losses.

Turning our attention to equities in the EEs, we notice that the required premium on the associated HML strategy has increased dramatically at the beginning of the pandemic and it has followed the contagion waives that we have observed over the last 20 months. The initial jump should not be surprising as both China and India are in the High-COVID portfolio. It is interesting, however, that the response to global news of High- and Low-COVID EEs quickly became less heterogeneous by the end of April. At the time we are writing this manuscript, our estimation suggests that the HML-COVID is quantitatively very similar across AE and EE equity markets.

Additional results with daily data. In table B.1 (see Appendix B), we show that replacing covid-related news with market returns in our conditional model delivers no positive and statistically significant market price of risk. This result confirms that (i) a conditional CAPM model fails in capturing viral contagion risk; and (ii) our measures are informative about viral risk.

So far, we have estimated a model with heterogeneous and time-varying exposure to a common risk factor related to global contagion news. Our dataset enables us also to construct AE- and EE-specific measures of both COVID19 case growth and Twitter tone. See, for example, figure B.1 in the Appendix.

We identify purely AE- and EE-specific components by regressing these fundamental measures on their global counterpart. The residuals of these two separate regressions represent for us AE- and EE-specific news. In Appendix B, table B.2, we show mixed results. Specifically, when we use only equity-based test assets, local contagion news (panel A) are priced negatively in AEs and positively in EEs. Twitter-based local news (panel B) have a market price of risk statistically not different from zero. Only when we use both bond and equity indices as test assets, local news are priced. Given these considerations, we consider our specification with heterogeneous and time-varying exposure to global contagion risk news as more robust.

Intra-day news. An important advantage of our Twitter-based risk-factor is that we can measure it at very high frequencies, in contrast to daily contagion cases. Using higher frequency data may help sharpen the estimate of the market price of risk because it provides an increased number of observations.

In this section, we focus only on European countries whose markets are open simultaneously. Specifically, we focus on ITA, ESP, UK, FRA, DEU, CHE, and SWE. Every day, we group them into three portfolios according to their relative number of COVID19 cases. In table 6, we show our estimation results when we link hourly equity and bond excess returns to hourly Twitter-based news.

As for daily data, we consider multiple specifications of our no-arbitrage model. In this case, we also report our estimated beta coefficients. The implied market price of risk is positive, well identified, and sizable. Our implied betas continue to be positive, i.e., viral contagion is priced as a source of risk. Consistent with the failure of the international-CAPM documented in table B.1,

Table 6. Hourly Conditional Linear Factor Model

	β_0	$\beta_{L,1}$	$\beta_{M,1}$	$\beta_{H,1}$	MPR	N.Obs	N. Assets
			Panel A: equities and bonds, equities betas				
Hourly log returns							
coef	-0.077^{***}	8.387^{***}	3.376^{***}	2.473^{***}	0.021^{***}	3286	6
se	(0.007)	(0.695)	(0.280)	(0.207)	(0.003)	3286	6
Hourly log EUR returns (adjusting for FX)							
coef	-0.081^{***}	8.830^{***}	3.577^{***}	2.623^{***}	0.021^{***}	3286	6
se	(0.006)	(0.669)	(0.272)	(0.202)	(0.003)	3286	6
Hourly log returns controlling for the Market							
coef	-0.064^{***}	6.599^{***}	2.754^{***}	2.098^{***}	0.022^{***}	3124	6
se	(0.007)	(0.751)	(0.315)	(0.241)	(0.003)	3124	6
			Panel B: equities and bonds, bond betas				
Hourly log returns							
coef	-0.053^{***}	5.864^{***}	2.303^{***}	1.703^{***}	0.021^{***}	3286	6
se	(0.005)	(0.497)	(0.196)	(0.147)	(0.003)	3286	6
Hourly log EUR returns (adjusting for FX)							
coef	-0.056^{***}	6.152^{***}	2.450^{***}	1.809^{***}	0.021^{***}	3286	6
se	(0.004)	(0.472)	(0.189)	(0.141)	(0.003)	3286	6
Hourly log returns controlling for the Market							
coef	-0.043^{***}	4.531^{***}	1.889^{***}	1.440^{***}	0.022^{***}	3124	6
se	(0.005)	(0.526)	(0.220)	(0.168)	(0.003)	3124	6

Notes: This table shows the results of the conditional linear factor model described in equations (2)–(4). Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). The coefficient $\beta_{f,t} = \beta_0 + \beta_f X_{f,t}$ refers to the exposure of the equity portfolio $f \in \{H, M, L\}$ to the COVID19 factor. We measure hourly COVID19 news as unexpected improvement in the hourly tone of COVID19-related tweets. Both hourly excess returns and market prices of risk are in log units. When we control for the market, returns are in USD, the market is measured by the MSCI Global Index and our factor model comprises a total of two factors. Our real-time data range from February 2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

our the implied market price of risk is still positive and sizable when we control for the market and use a broader cross section of test assets.

Controlling for Volatility. In this last step of our research, we project our Tweeter-based COVID factor on realized market volatility and use the implied residual to redo our analysis. Equivalently, we look at COVID news that are orthogonal to pure volatility shocks. We measure realized volatility as the standard deviation of the MSCI Global Index at the daily (hourly) frequency using

Table 7. Vol-Adjusted Conditional Linear Factor Model

	β_0	$\beta_{L,1}$	$\beta_{M,1}$	$\beta_{H,1}$	MPR	N.Obs	N. Assets
			Panel A: equities, news from Twitter				
Daily log returns							
coef	-0.271^{***}	40.368^{***}	19.288^{***}	5.722^{***}	0.012^{***}	412	3
se	(0.051)	(5.972)	(2.643)	(1.565)	(0.003)	412	3
			Panel B: equities, news from Twitter				
Hourly log returns							
coef	-0.029^{***}	2.652^{***}	1.078^{***}	0.881^{***}	0.014^{***}	3444	6
se	(0.002)	(0.240)	(0.102)	(0.078)	(0.005)	3444	6

Notes: This table shows the results of the conditional linear factor model described in equations (2)–(4). Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). The coefficient $\beta_{f,t} = \beta_0 + \beta_f X_{f,t}$ refers to the exposure of the equity portfolio $f \in \{H, M, L\}$ to the COVID19 factor. We measure hourly (daily) COVID19 news as unexpected improvement in the hourly (daily) tone of COVID19-related tweets. We project this factor on realized market volatility and use the implied residual in our estimation. Both excess returns and market prices of risk are in log units and are expressed in USD. The market is measured by the MSCI Global Index. Our real-time data range from February 2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

a rolling window of a trading week (a single trading day). We report our results in table 7. Both daily data and intra-day data confirm that contagion news have an extremely high MPR, even after controlling for volatility.

International Flows. In order to further validate our results, we study international investment flows related to the countries in our cross section. Weekly net flows are from EPFR and they are rescaled by country-level GDP so that our results are not driven by country size. In this step, we exclude the US given its special role played in international markets (among others, see Maggiori 2017). After forming portfolios according to relative contagion levels, we forecast one-week ahead flows using the (lagged) weekly share of portfolio-level COVID19 cases. As reported in table 8, countries that start the week with a higher level of relative contagion are expected to receive lower net inflows ($\beta_1 < 0$)). This effect is reversed ($\beta_1 > 0$)) when we focus on net bond flows in EE, consistent with the idea that they may be perceived as safer assets and hence their demand may actually increase due to flight to safety. As shown in figure 13, low-COVID countries tend to

Table 8. International Flows and News

	Bonds		Equities	
	AE	EE	AE	EE
β_0	0.247***	−0.171***	0.589***	0.135***
	(0.026)	(0.032)	(0.051)	(0.036)
β_1	−0.921***	0.311***	−4.500***	−0.996***
	(0.073)	(0.070)	(0.274)	(0.123)
J-stat	11.234	11.825	7.069	11.676
N	75	71	75	71

Notes: This table reports the results of the following linear system:

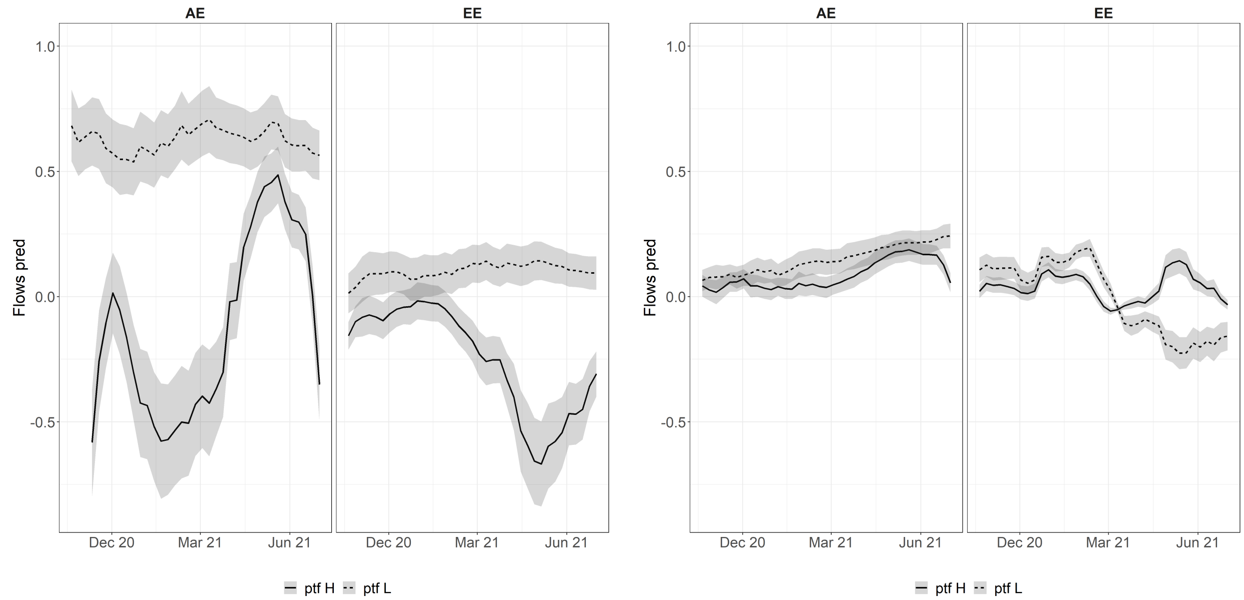
$$FL_t^f = \beta_0 + \beta_1 X_{t-1}^f + \epsilon_t^f$$

where FL_t^f is the flow to funds that invest in portfolio $f \in \{H, M, L\}$ during week t rescaled by portfolio$-f$ 2019 GDP; X_{t-1}^f refers to the weekly share of portfolio-specific COVID19 cases. Portfolios are formed on a weekly basis according to the relative share of country-specific COVID19 cases measured the week before formation. Fund flows-to-GDP is expressed in basis points (bps). Our data range from February 2020 to the date of this manuscript at a weekly frequency. Estimates and HAC-adjusted standard errors are obtained through GMM.

receive a higher net inflow than high-COVID countries. This statement, however, does not apply to bonds in EEs. During the summer 2021, high-covid EEs have experienced higher inflows for their sovereign bonds.

4 Conclusion

In this study, we quantify the exposure of major financial markets to news shocks about global contagion risk while accounting for local epidemic conditions. We construct a novel data set comprising (i) medical announcements related to COVID19 for a wide cross section of countries; and (ii) high-frequency data on epidemic news diffused through Twitter. Across several classes of financial assets and currencies, we provide novel empirical evidence about financial dynamics (i) around epidemic announcements, (ii) at a daily frequency, and (iii) at an intra-daily frequency. Formal estimations based on both contagion data and social media activity about COVID19 confirm that the market price of epidemic risk is very significant. In the spirit of Mulligan (2020), we conclude



Equities *Bonds*

Fig. 13. Expected Investment Flows

Notes: For each asset class, we depict forecasted net investment flows. 'ptf H' ('ptf L') refers to a portfolio of countries with relatively high (low) contagion cases. We split our sample across advanced and emerging economies (AE and EE, respectively). The estimates are based on the following linear system:

$$FL_t^f = \beta_0 + \beta_1 X_{t-1}^f + \epsilon_t^f$$

where FL_t^f is the flow to funds that invest in portfolio $f \in \{H, M, L\}$ during week t rescaled by portfolio$-f$ 2019 GDP; X_{t-1}^f refers to the weekly share of portfolio-specific COVID19 cases. Portfolios are formed on a weekly basis according to the relative share of country-specific COVID19 cases measured the week before formation. Fund flows-to-GDP is expressed in basis points (bps). Our data range from February 2020 to the date of this manuscript at a weekly frequency. Estimates and HAC-adjusted standard errors are obtained through GMM.

that policies related to prevention and containment of contagion could be very valuable not only in terms of lives saved but also in terms of global wealth.

References

Acemoglu, D., V. Chernozhukov, I. Werning, and M. D. Whinston. 2020. Ta multi-risk sir model with optimally targeted lockdown. *NBER Working Paper No. 27102.*

Acharya, V. V., and S. Steffen. 2020. The risk of being a fallen angel and the corporate dash for cash in the midst of covid. *CEPR's Covid Economics Review No. 10.*

Ai, H., and R. Bansal. 2018. Risk preferences and the macroeconomic announcement premium. *Econometrica* 86(4).

Albuquerque, R. A., Y. J. Koskinen, S. Yang, and C. Zhang. 2020. Love in the time of covid-19: The resiliency of environmental and social stocks. *CEPR's Covid Economics Review No. 11.*

Alfaro, L., A. Chari, A. N. Greenland, and P. K. Schott. 2020. Aggregate and firm-level stock returns during pandemics, in real time. *CEPR's Covid Economics Review No. 4.*

Alon, T. M., M. Doepke, J. Olmstead-Rumsey, and M. Tertilt. 2020. The impact of covid-19 on gender equality. *CEPR's Covid Economics Review No. 4.*

Alvarez, F. E., D. Argente, and F. Lippi. A simple planning problem for covid-19 lockdown. Working Paper 26981 National Bureau of Economic Research April 2020 http://www.nber.org/papers/w26981.

Atkeson, A. How deadly is covid-19? understanding the difficulties with estimation of its fatality rate. Working Paper 26965 National Bureau of Economic Research April 2020 http://www.nber.org/papers/w26965.

Augustin, P., V. Sokolovski, M. G. Subrahmanyam, and D. Tomio. 2021. In sickness and in debt: The covid-19 impact on sovereign credit risk. *Journal of Financial Economics.*

Bahaj, S., and R. Reis. 2020. Central bank swap lines during the covid-19 pandemic. *CEPR's Covid Economics Review No. 4.*

Baker, S. R., N. Bloom, and S. J. Davis. 2016. Measuring economic policy uncertainty*. *The Quarterly Journal of Economics* 131(4):1593.

Baker, S. R., N. Bloom, S. J. Davis, K. Kost, M. Sammon, and T. Viratyosin. 2020a. The unprecedented stock market reaction to covid-19. *The Review of Asset Pricing Studies* 10(4).

Baker, S. R., N. Bloom, S. J. Davis, and S. J. Terry. Covid-induced economic uncertainty. Working Paper 26983 National Bureau of Economic Research April 2020b http://www.nber.org/papers/w26983.

Barro, R. J., J. F. Ursúa, and J. Weng. The coronavirus and the great influenza pandemic: Lessons from the "spanish flu" for the coronavirus's potential effects on mortality and economic activity. Working Paper 26866 National Bureau of Economic Research March 2020 http://www.nber.org/papers/w26866.

Barrot, J.-N., B. Grassi, and J. Sauvagnat. 2020. Sectoral effects of social distancing. *CEPR's Covid Economics Review No. 3*.

Bartik, A. W., M. Bertrand, Z. B. Cullen, E. L. Glaeser, M. Luca, and C. T. Stanton. How are small businesses adjusting to covid-19? early evidence from a survey. Working Paper 26989 National Bureau of Economic Research April 2020 http://www.nber.org/papers/w26989.

Bianchi, F., R. Cram, and F. Kung. 2021. Using social media to identify the effects of congressional viewpoints on asset prices. *Working Paper*.

Bonaccolto, G., N. Borri, and A. Consiglio. 2019. Breakup and default risks in the eurozone. *Available at SSRN 3487453*.

Bordalo, P., N. Gennaioli, Y. Ma, and A. Shleifer. 2020. Overreaction in macroeconomic expectations. *American Economic Review* 110(9).

Bretscher, L., A. Hsu, P. Simasek, and A. Tamoni. 2020a. Covid-19 and the cross-section of equity returns: Impact and transmission. *The Review of Asset Pricing Studies* 10(4).

Bretscher, L., A. Hsu, and A. Tamoni. 2020b. The supply channel of uncertainty shocks and the cross-section of returns: Evidence from the covid-19 crisis. *Available at SSRN 3588418*.

Bybee, L., B. T. Kelly, A. Manela, and D. Xiu. 2020. The structure of economic news.

Caballero, R. J., and A. Simsek. 2020. A model of asset price spirals and aggregate demand amplification of a "covid-19" shock. *CEPR Discussion Paper*.

Calomiris, C. W., and H. Mamaysky. 2019. How news and its context drive risk and returns around the world. *Journal of Financial Economics* 133(2):299–336.

Carletti, E., T. Oliviero, M. Pagano, L. Pelizzon, and M. G. Subrahmanyam. 2020. The covid-19 shock and equity shortfall: Firm-level evidence from italy.

Chiou, L., and C. Tucker. Social distancing, internet access and inequality. Working Paper 26982 National Bureau of Economic Research April 2020 http://www.nber.org/papers/w26982.

Coibion, O., Y. Gorodnichenko, and M. Weber. 2020. The cost of the covid-19 crisis: Lockdowns, macroeconomic expectations, and consumer spending. *BFI Working Paper.*

Cookson, J. A., J. Engelberg, and W. Mullins. 2021. Echo chambers. *Working paper.*

Cororaton, A., and S. Rosen. 2020. Public firm borrowers of the us paycheck protection program. *CEPR's Covid Economics Review No. 15.*

Correia, S., S. Luck, and E. Verner. 2020. Pandemics depress the economy, public health interventions do not: Evidence from the 1918 flu. *CEPR working paper.*

Corsetti, G., M. Bodenstein, and L. Guerrieri. 2020. Social distancing and supply disruptions in a pandemic. *CEPR Discussion Paper.*

Coven, J., and A. Gupta. 2020. Disparities in mobility response to covid-19. *Working paper.*

Cox, J., , D. L. Greenwald, and S. C. Ludvigson. 2020. What explains the covid-19 stock market? *Working Paper.*

Da, Z., J. Engelberg, and P. Gao. https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1540-6261.2011.01679.x 2011. In search of attention. *The Journal of Finance* 66(5):1461–1499. doi: 10.1111/j.1540-6261.2011.01679.x.

Eichenbaum, M. S., S. Rebelo, and M. Trabandt. The macroeconomics of epidemics. Working Paper 26882 National Bureau of Economic Research March 2020 http://www.nber.org/papers/w26882.

Elgin, C., G. Basbug, and A. Yalaman. 2020. Economic policy responses to a pandemic: Developing the covid-19 economic stimulus index. *CEPR's Covid Economics Review No. 3.*

Engle, R. F., S. Giglio, B. Kelly, H. Lee, and J. Stroebel. 2020. Hedging Climate Change News. *The Review of Financial Studies* 33(3).

Farboodi, M., G. Jarosch, and R. Shimer. 2020. Internal and external effects of social distancing in a pandemic. *CEPR's Covid Economics Review No. 9.*

Faria-e Castro, M., and F. S. Louis. 2020. Fiscal policy during a pandemic. *Federal Reserve Bank of St. Louis.* Working paper.

Favero, C. A. 2020. Why is covid-19 mortality in lombardy so high? evidence from the simulation of a seihcr model. *CEPR's Covid Economics Review No. 4.*

Fornaro, L., and M. Wolf. 2020. Covid-19 coronavirus and macroeconomic policy. *CEPR working paper.*

Garmaise, M., Y. Levi, and H. Lustig. 2021. Spending less after (seemingly) bad news. *Working paper.*

Gerding, F., T. Martin, and F. Nagler. April 2020. The value of fiscal capacity in the face of a rare disaster. *Working paper.*

Giglio, S., M. Maggiori, J. Stroebel, and S. Utkus. 2020. Inside the mind of a stock market crash. *arXiv preprint arXiv:2004.01831.*

Glover, A., J. Heathcote, D. Krueger, and J.-V. Rios-Rull. 2020. Health versus wealth: On the distributional effects of controlling a pandemic. *CEPR's Covid Economics Review No. 6.*

Goldstein, I., R. Koijen, and H. Mueller. 2021. Covid-19 and its impact on financial markets and the real economy. *The Review of Financial Studies.*

Gormsen, N. J., and R. S. J. Koijen. 2020. Coronavirus: Impact on stock prices and growth expectations. *The Review of Asset Pricing Studies* 10(4):574–597.

Gormsen, N. J., R. Koijen, and I. Martin. 2021. Implied dividend volatility and expected growth. *Working Paper.*

Hagedorn, M., and K. Mitman. 2020. Corona policy according to hank. *CEPR Discussion Paper.*

Han, L. J. 2020. Iannouncements, expectations, and stock returns with asymmetric information. *Working Paper.*

Hartley, J. S., and A. Rebucci. 2020. An event study of covid-19 central bank quantitative easing in advanced and emerging economies.

Hassan, T. A., S. Hollander, L. Van Lent, and A. Tahoun. 2019. Firm-level political risk: Measurement and effects. *The Quarterly Journal of Economics* 134(4):2135–2202.

Hassan, T. A., S. Hollander, L. van Lent, and A. Tahoun. 2020. Firm-level exposure to epidemic diseases: Covid-19, sars, and h1n1. *Working Paper Series.*

Hensvik, L., T. Le Barbanchon, and R. Rathelot. 2020. Which jobs are done from home? evidence from the american time use survey. *CEPR Discussion Papers.*

Ichino, A., C. A. Favero, and A. Rustichini. 2020. Restarting the economy while saving lives under covid-19. *CEPR Discussion Papers.*

Israelsen, R., N. Gupta, and G. Eitan. 2021. Political polarization in financial news. *Working paper.*

Jones, C. J., T. Philippon, and V. Venkateswaran. Optimal mitigation policies in a pandemic: Social distancing and working from home. Working Paper 26984 National Bureau of Economic Research April 2020 http://www.nber.org/papers/w26984.

Kaniel, R., and P. Wang. 2020. Unmasking mutual fund derivative use during the covid-19 crisis. *CEPR's Covid Economics Review No. 49.*

Krueger, D., H. Uhlig, and T. Xie. 2020. Macroeconomic dynamics and reallocation in an epidemic. *CEPR's Covid Economics Review No. 5.*

Lewis, D., K. Mertens, and J. H. Stock. 2020. U.s. economic activity during the early weeks of the sars-cov-2 outbreak. *CEPR's Covid Economics Review No. 6.*

Lucca, D. O., and E. Moench. 2015. The pre-fomc announcement drift. *The Journal of Finance* 70(1).

Ludvigson, S. C., S. Ma, and S. Ng. 2020. Covid19 and the macroeconomic effects of costly disasters. *CEPR's Covid Economics Review No. 9.*

Ma, C., J. H. Rogers, and S. Zhou. 2020. Global economic and financial effects of 21st century pandemics and epidemics. *CEPR's Covid Economics Review No.5.*

Maggiori, M. 2017. Financial Intermediation, International Risk Sharing, and Reserve Currencies. *American Economic Review* 107(10).

Manela, A., and A. Moreira. 2017. News implied volatility and disaster concerns. *Journal of Financial Economics* 123(1):137–162.

Mulligan, C. B. Economic activity and the value of medical innovation during a pandemic. Working Paper 27060 National Bureau of Economic Research April 2020 http://www.nber.org/papers/w27060.

Papanikolaou, D., and L. D. W. Schmidt. 2020. Working remotely and the supply-side impact of covid-19. *Working Paper.*

Pástor, L., and B. Vorsatz. 2020. Mutual fund performance and flows during the covid-19 crisis. *The Review of Asset Pricing Studies* 10(4).

Pelger, M. 2020. Understanding systematic risk - a high-frequency approach. *The Journal of Finance.*

Ramelli, S., and A. Wagner. 2020. Feverish stock price reactions to covid-19. *Forthcoming, Review of Corporate Finance Studies; Swiss Finance Institute Research Paper No. 20-12.*

Schoenfeld, J. 2020. The invisible risk: Pandemics and the financial markets. *Available at SSRN 3567249.*

Sinagl, P. 2020. Cash-flow risk and industry response to covid-19 outbreak. *Available at SSRN 3566511.*

Tetlock, P. C. 2007. Giving content to investor sentiment: The role of media in the stock market. *The Journal of finance* 62(3):1139–1168.

Twedt, B., and L. Rees. 2012. Reading between the lines: An empirical examination of qualitative attributes of financial analysts' reports. *Journal of Accounting and Public Policy* 31(1):1–21.

Appendix A. Data Sources

Table A.1: **News Papers**

Country	**Newspaper**	**Twitter Account**	**BBD**	**Language**
Argentina	La Nacion	@LANACION		Spanish
Argentina	Clarin	@clarincom		Spanish
Argentina	Diario Cronica	@cronica		Spanish
Argentina	Infobae	@infobae		Spanish
Australia	The Age	@theage		English
Australia	The Australian	@australian		English
Australia	The Daily Telegraph	@dailytelegraph		English
Australia	Financial Review	@FinancialReview		English
Brazil	O Globo	@JornalOGlobo		Portuguese
Brazil	O Estado de Sao Paulo	@Estadao		Portuguese
Brazil	Folha de S.Paulo	@folha		Portuguese
Brazil	Gaucha ZH	@GauchaZH		Portuguese
Canada	Gazette	@mtlgazette	Yes	English
Canada	Globe and Mail	@globeandmail	Yes	English
Canada	Ottawa Citizen	@OttawaCitizen	Yes	English
Canada	Toronto Star	@TorontoStar	Yes	English
Canada	Vancouver Sun	@VancouverSun	Yes	English
Chile	La Tercera	@latercera		Spanish
Chile	BioBioChile	@biobio		Spanish
Chile	El Mostrador	@elmostrador		Spanish
Chile	The Clinic	@theclinicc1		Spanish
China	People's Daily, China	@PDChina		English

(To be continued)

Country	Newspaper	Twitter Account	BBD	Language
China	China Xinhua News	@XHNews		English
China	China Daily	@ChinaDaily		English
Colombia	El Espectador	@elespectador		Spanish
Colombia	El Colombiano	@elcolombiano		Spanish
Colombia	El Heraldo	@elheraldoco		Spanish
Colombia	El Tiempo	@ELTIEMPO		Spanish
France	Le Monde	@lemondefr	Yes	French
France	Le Figaro	@Le_Figaro		French
France	Liberation	@libe		French
France	Le Parisien	@le_Parisien		French
Germany	Handelsblatt	@handelsblatt	Yes	German
Germany	Frankfurter Allgemeine Zeitun	@faznet	Yes	German
Germany	BILD	@BILD		German
Germany	Zeit Online	@zeitonline		German
Hong Kong	South China Morning Post	@SCMPNews	Yes	English
Hong Kong	Hong Kong Free Press	@HongKongFP		English
Hong Kong	RTHK English News	@rthk_enews		English
India	Economic Times	@EconomicTimes	Yes	English
India	Times of India	@timesofindia	Yes	English
India	Hindustan Times	@htTweets	Yes	English
India	The Hindu	@the_hindu	Yes	English
Italy	Corriere Della Sera	@Corriere	Yes	Italian
Italy	La Repubblica	@repubblica	Yes	Italian
Italy	Il Sole 24 ORE	@sole24ore		Italian
Japan	Asahi Shimbun AJW	@AJWasahi	Yes	English

(To be continued)

Country	**Newspaper**	**Twitter Account**	**BBD**	**Language**
Japan	The Japan News by Yomiuri	@The_Japan_News	Yes	English
Japan	The Japan Times	@japantimes		English
Japan	Japan Today News	@JapanToday		English
Korea	Korea JoongAng Daily	@JoongAngDaily		English
Korea	The Korea Herald	@TheKoreaHerald		English
Korea	Yonhap News Agency	@YonhapNews		Korean
Korea	The Korea Times	@koreatimescokr		Korean
Mexico	La Jornada	@lajornadaonline		Spanish
Mexico	Reforma	@Reforma		Spanish
Mexico	El Universal	@El_Universal_Mx		Spanish
Mexico	Milenio	@Milenio		Spanish
New Zealand	The New Zealand Herald	@nzherald		English
New Zealand	The Sydney Morning Herald	@smh		English
New Zealand	Herald Sun	@theheraldsun		English
New Zealand	Guardian Australia	@GuardianAus		English
Spain	EL MUNDO	@elmundoes	Yes	Spanish
Spain	EL PAIS	@el_pais	Yes	Spanish
Spain	ABC.es	@abc_es		Spanish
Spain	La Vanguardia	@LaVanguardia		Spanish
Switzerland	Neue Zurcher Zeitung	@NZZ		German
Switzerland	20 Minuten	@20min		German
Switzerland	24heures	@24heuresch		French
Switzerland	Le Temps	@LeTemps		French
USA	LA Times	@latimes	Yes	English
USA	USA Today	@USATODAY	Yes	English

(To be continued)

Country	**Newspaper**	**Twitter Account**	**BBD**	**Language**
USA	Chicago Tribune	@chicagotribune	Yes	English
USA	Washinton Post	@washingtonpost	Yes	English
USA	Boston Globe	@BostonGlobe	Yes	English
USA	Wall Street Journal	@WSJ	Yes	English
USA	Miami Herald	@MiamiHerald	Yes	English
USA	Dallas Morning News	@dallasnews	Yes	English
USA	Houston Chronicle	@HoustonChron	Yes	English
USA	San Fransisco Chronicle	@sfchronicle	Yes	English
USA	New York Times	@nytimes	Yes	English
UK	The Times	@thetimes	Yes	English
UK	Financial Times	@FinancialTimes	Yes	English
UK	BBC News (UK)	@BBCNews		English
UK	Guardian news	@guardiannews		English

Notes: This table reports our newspaper sources. For each newspaper, we specify headquarter location, original language, and twitter account. A 'Yes' under the column BBD denotes a newspaper used also in Baker et al. (2016).

Table A.2. Computing Tone of Tweets: Two Examples

Tweet Text	**Positive Words**	**Negative Words**	**Tone**
The coronavirus pandemic has been particularly devastating to the United States's biggest cities. It comes as the country's major urban centers were already losing their appeal for many Americans.	"devastating", "losing"	"appeal"	$\frac{-2+1}{3} = -0.33$
A shortage of test kits and technical flaws in the U.S. significantly delayed widespread coronavirus testing. This is how testing has increased since the beginning of March — and how far it still needs to go, according to the Harvard estimates	"shortag", "flaws", "delayed"		$\frac{-3}{3} = -1$

Notes: This table shows two examples of the computation of the tone of a tweet using Polyglot.

Table A.3. Data Sources

Country	Equity Index	Equity Volume Index	Long Term Bond Index	Sovereign CDS	Short Term Bond Index	Currency
Canada	SPTSX Composite Index	TSXVOL Index	GCAN10YR INDEX	CAGV5YUSAC	CA 3M benchmark rate	USDCAD
China	SHSZ300 INDEX	SHSZ300V INDEX	GCNY10YR INDEX	CNGV5YUSAC	CN 1Y benchmark rate	USDCNY
France	CAC Index	CACVOLC Index	GECU10YR INDEX	FRGV5YUSAC	FR 3M benchmark rate	EURUSD
Germany	DAX Index	DAXVOLC Index	GDBR10 INDEX	DEGV5YUSA	DE 3M benchmark rate	EURUSD
Hong Kong	HSI INDEX	HSIVOLC INDEX	HKGG10Y Index	HKGV5YUSAC	HK 3M benchmark rate	USDHKD
Italy	FTSE MIB Index	FTMIBVOL Index	GBTPGR10 INDEX	ITGV5YUSAC	IT 3M benchmark rate	EURUSD
India	SENSEX INDEX	SNSXVOLC INDEX	GIND10YR INDEX	INGV5YUSAC	ES 3M benchmark rate	USDINR
Japan	NKY INDEX	NKYVOLC INDEX	GJGB10 INDEX	JPGV5YUSAC	JP 3M benchmark rate	USDJPY
Korea	KOSPI Index	KOSPIVOLC INDEX	GVSK10YR INDEX	KRGV5YUSAC	KR 1Y benchmark rate	USDKRW
New Zealand	NZSE50FG INDEX	NZ50VOL Index	GNZGB10 INDEX	NZGV5YUSAQ	NZ 3M benchmark rate	NZDUSD
Spain	IBEX 35	IBEXVOLC INDEX	GSPG10YR INDEX	ESGV5YUSAC	ES 3M benchmark rate	EURUSD
Switzerland	SMI Index	SMIVOLC Index	GSWISS10 INDE	CHGV5YUSAC	CH 3M benchmark rate	USDCHF
Sweden	OMXS30 Index	OMXVOLC Index	GSGB10YR INDEX	SEGV5YUSAC	SE 3M benchmark rate	USDSEK
USA	SPX Index	SPXVOLC Index	USGG10YR INDEX	USGV5YEUAC	US 3M benchmark rate	USD
UK	UKX INDEX	UKXVOLC INDEX	GUKG10 INDEX	GBGV5YUSAC	GB 3M benchmark rate	GBPUSD
Source	Bloomberg	Bloomberg	Bloomberg	Thomson Reuters	Bloomberg	Bloomberg
Frequency	Minute	Minute	Minute	Day	Minute	Minute

Notes: This table shows our data sources.

Appendix B. Additional Estimation Results

Table B.1. Summary of MPR Estimation: Conditional CAPM

	Equity		Bonds & Equity	
	A.E.	E.E.	A.E.	E.E.
Local units				
coef	-0.017^{***}	0.005	-0.017^{***}	0.004
se	(0.005)	(0.005)	(0.002)	(0.003)
USD units				
coef	-0.024^{***}	0.009	-0.024^{***}	0.003
se	(0.006)	(0.005)	(0.004)	(0.003)

Notes: This table shows the results of the conditional linear factor model described in equations (2)–(4) where the risk factor is measured by the news in the MSCI Global Index. Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). Both daily excess returns and market prices of risk are in log units and expressed in USD. Our real-time data range from 2/1/2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

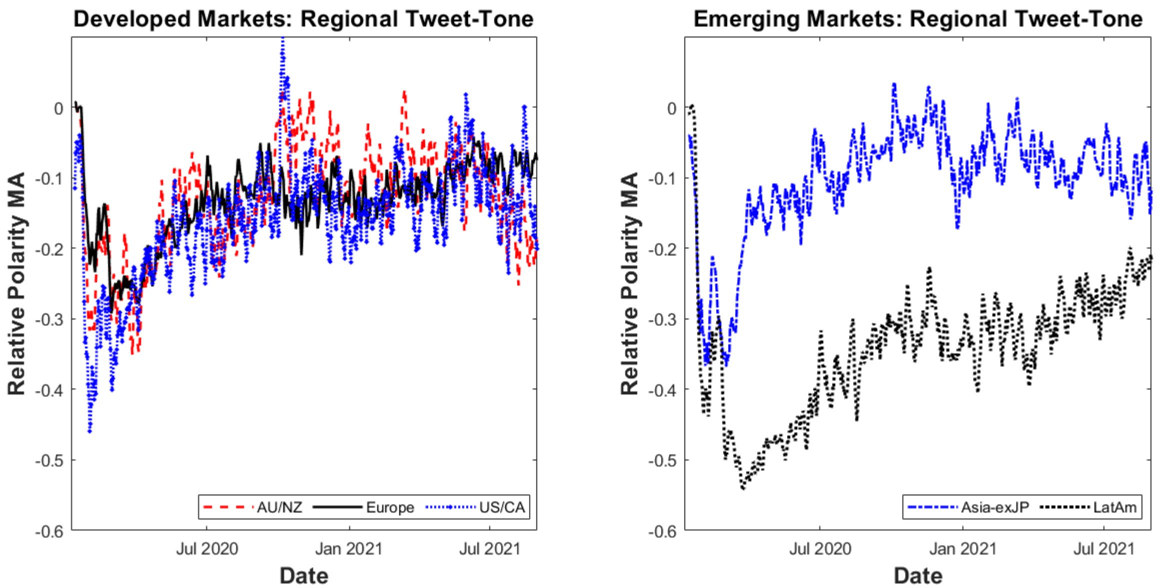


Fig. B.1. Regional Twitter-Based Tone

Notes: This figure shows our daily Twitter-based tone for different countries. We use Polygot to measure the polarity of our tweets and compute the tone of each tweet according to Twedt and Rees (2012). We aggregate the tones at a daily frequency and across regions. MA refers to a backward looking 5-day moving average.

Table B.2. Summary of MPR Estimation: Local News

	Equity		Bonds & Equity	
	A.E.	E.E.	A.E.	E.E.
	Panel A: Local News about Covid cases			
Local units				
coef	-0.007^{***}	0.009^{***}	-0.003^{***}	-0.005^{***}
se	(0.002)	(0.001)	(0.001)	(0.001)
USD units				
coef	-0.007^{***}	0.006^{***}	-0.004^{***}	-0.001
se	(0.003)	(0.001)	(0.001)	(0.002)
Controlling for MKT				
coef	0.000	0.002	-0.003^{***}	-0.007^{***}
se	(0.001)	(0.002)	(0.001)	(0.001)
	Panel B: Local News from Twitter			
Local units				
coef	0.032	0.283	0.027^{***}	0.007^{***}
se	(0.021)	(0.414)	(0.005)	(0.001)
USD units				
coef	0.036	-0.013	0.015^{***}	0.007^{***}
se	(0.029)	(0.022)	(0.003)	(0.001)
Controlling for MKT				
coef	0.004	0.001	0.008^{***}	0.006^{***}
se	(0.003)	(0.003)	(0.002)	(0.001)

Notes: This table shows the results of the conditional linear factor model described in equations (2)–(4) applied to AE- and EE-specific news. Portfolios are formed on a daily basis according to the relative share of country-specific COVID19 cases measured the day before formation (X_t). In panel A (panel B), the COVID19 factor is measured as the news to local COVID cases growth (tone of COVID-related tweets). When we measure the COVID19 news as unexpected number of contagion cases (unexpected improvement in COVID19-related tweets), we expect a negative (positive) market price of risk (MPR). Both daily excess returns and market prices of risk are in log units. The last two columns are based on a broader cross section of test assets comprising both equity and bond portfolios. When we control for the market, returns are in USD, the market is measured by the MSCI Global Index, and our factor model comprises a total of two factors. Our real-time data range from 2/1/2020 to the date of this manuscript. Estimates and HAC-adjusted standard errors are obtained through GMM.

Chapter 3. Cover your assets: non-performing loans and coverage ratios in Europe[1]

1. Introduction

One of the most debated issues in Europe since the financial and sovereign debt crises concerns the accumulation of large stocks of non-performing loans (NPLs) and the numerous policy actions put forth to deal with this problem. Unfortunately, despite all the undertaken efforts, because of the COVID-19 pandemic and the associated economic recession – far worse than that triggered by the global financial crisis – the issue of surging NPL stocks is bound to be a policy priority once again (Ari et al., 2020).

European policymakers have faced and studied in detail the consequences of high volumes of NPLs, namely limited bank lending ability, impairment of the monetary policy mechanism, and reduced output growth (Draghi, 2017; ESRB, 2019). In designing measures to limit the consequences of high NPLs, particular attention has been dedicated to banks' provisioning and loss coverage policies. In fact, while high amounts of NPLs are certainly problematic, the level of loss coverage, i.e., the amount of loan loss reserves (LLRs), determines how losses originating from NPLs impact bank capital (Constâncio, 2017). To explain the mechanism, each year banks set aside loan loss provisions (LLPs), to form loan loss reserves. These reserves work as a buffer to absorb the expected loan losses because, when the loss occurs, banks can draw on these reserves without impairing their capital. Hence, it is not the amount of NPLs *per se*, but the "uncovered" portion of NPLs that represents the real threat to bank balance sheets.

Against this background, numerous policy initiatives have been adopted to enhance banks' coverage policies and, specifically, to increase the coverage ratio (i.e., the share of loan loss reserves over NPLs), which has gained relevance as a key prudential tool and supervisory metric of bank soundness (ECB, 2016 and 2017a). *Ceteris paribus*, banks with larger volumes of NPLs and lower coverage ratios are more

[1] This chapter is a joint work with Lucia Alessi, Brunella Bruno, Elena Carletti and Katja Neugebauer.
Lucia Alessi is affiliated with the European Commission, Joint Research Centre, European Commission, email: lucia.alessi@ec.europa.eu. Brunella Bruno is affiliated with Università Luigi Bocconi, email: brunella.bruno@unibocconi.it; Elena Carletti is affiliated with Università Luigi Bocconi, IGIER and CEPR, email: elena.carletti@unibocconi.it; Katja Neugebauer is affiliated with Banco de Portugal, Financial Stability Department, email: kneugebauer@bportugal.pt.

vulnerable to negative shocks affecting borrowers' credit quality, especially in bad times, when loan losses are more likely.[2] It follows that in a situation where NPLs are bound to increase, banks should react promptly to preserve an adequate loss coverage. This is for example what is happening in the COVID-19 crisis, where some large banks have started accumulating provisions in anticipation of future losses on their stocks of loans.

Nevertheless, loan loss coverage policies still vary largely across banks and countries in Europe, with many of the countries with the highest level of NPLs reporting below-average coverage ratios (EBA, 2018). [3] In this paper, we exploit this variation to investigate drivers and dynamics of bank coverage ratios and their components at both the micro (bank) and macro (country) level, using a sample of around 440 large and medium-sized banks in Europe over the period 2010–2017. The focus on Europe provides an interesting case study, given the high level of NPLs and the substantial bank and country heterogeneity in the region (EBA, 2018).

Our results point to the following three main conclusions. First, bank-specific factors are the main drivers of coverage ratios. This finding emphasizes the importance of micro prudential oversight as a way to induce banks to increase their coverage ratios. Still, some of the variation in coverage ratios is explained by unobservable, structural bank characteristics that could be better captured by close and customized scrutiny as it occurs in the supervisory dialogue. Among the bank-specific determinants, credit risk related factors such as reserve policies, (the level and change of) NPLs, credit growth, as well as forward-looking measures of credit risk play an important role. These results suggest that coverage ratios work more as a prudential (forward-looking) buffer than merely (and backward-looking) a booking account, even in a context where the "incurred loss" model (ILM) for calculating bank provisions is prevalent. This provides evidence of prudent behavior in setting coverage ratios even before the new accounting standard IFRS 9 was put into practice (as a forward-looking and ideally countercyclical approach to calculate provisions, IFRS 9 should lead to higher coverage ratios by promoting a timelier and more prudent provisioning).

Capitalization and cost efficiency also explain coverage policy variation, although to a lesser extent. In particular, increases in capitalization help banks enhance coverage ratios, i.e., one buffer

[2] Estimates report that net present value of NPLs may be as may be as low as 40-50% of the loan gross book value. Balance sheets are protected, and capital buffers remain impaired, as long as coverage ratios reflect this haircut (Fell et al., 2016).

[3] For example, large institutions have commonly reported lower coverage ratios than small and medium-sized banks. At the country level, the average coverage ratio in Europe is nearly 46%, but it ranges from 24% in Finland to nearly 70% in Hungary (EBA, 2018).

reinforces the other. Or, capital contraction gives banks an incentive to under-reserve, possibly to limit the immediate negative implications of higher provisioning on equity. Finally, an increase in the cost-to-income ratio is also associated with lower coverage level, possibly via incentives to under-reserve as in Ristolainen (2018).

The second main conclusion is that variations in NPLs or loan loss reserves affect coverage ratios in a non-trivial manner. In particular, by inspecting the underlying mechanisms, we show that when NPLs increase, banks tend to set aside larger reserves, but in a way that is not sufficient, at least in the short term, to determine higher coverage ratios. The relationship between coverage ratios and asset quality is, however, non-linear, as very high-NPL banks tend to be comparatively better covered as their asset quality worsens.[4] Moreover, by looking at well-reserved banks and risky banks (those with structurally high levels of LLRs and NPLs, respectively), we show that the former tend to have lower coverage ratios than the average bank, while the latter tend to be better covered. These findings altogether emphasize the need to look at reserves or the stock of NPLs only in conjunction with the associated level of coverage, as a comprehensive measure of balance sheet strength.

The third main conclusion concerns the effectiveness of a set of macro policies and policy tools in shaping coverage ratios, although as a less powerful alternative to micro supervision. In particular, we find that more stringent macroprudential policies (especially time-varying/dynamic loan-loss provisioning) are associated with higher coverage ratios. In addition, banks from high-NPL countries exhibit lower NPLs and lower coverage ratios in the presence of a better rule of law. This may suggest that stronger contract enforcement or more efficient courts support NPL resolution and thus decrease the need of large coverage. We also find that tighter capital rules are associated with lower NPLs and coverage ratios, but only in high-NPL countries, which call for a different calibration of such policies in different jurisdictions. This result is in line with the finding in Gropp et al. (2019) that banks tend to de-risk and deleverage in an attempt to comply with more stringent capital regulation.

Finally, we find higher coverage ratios in banks located in countries where secondary markets for distressed debt are larger, and even more so in banks located in high-NPL countries. This result corroborates the statements by European central authorities about the need to report adequate coverage

[4] As it is quite standard, also in supervisory reports, we measure asset quality with the level of NPLs. We are, however, aware that low NPLs do not necessarily translate in high quality of the underlying assets. One reason is that economic booms help loans remain performing. Another reason may also be that NPLs are low because of managerial under-reporting.

ratios to make loan disposals more likely and limit actual losses for the seller (Fell et al., 2016; Constâncio, 2017).

This paper also contributes to the literature on NPLs and provisioning. Despite the increased policy relevance, the empirical evidence on coverage ratios and its determinants remains scarce. Previous works on related topics have focused on explaining either NPLs or provisions, which, however, are rather uncorrelated with coverage ratios (see Table 1).[5] In fact, we find this ratio does not always move in the same direction of each of its component. In particular, unlike previous work on bank provisioning (see Laeven and Majoni, 2003, among others), managerial discretion to e.g., smooth earnings does not explain coverage policy, which instead responds primarily to non-discretionary factors related to expected credit risk. It follows that the coverage ratio is a more comprehensive indicator of balance-sheet strength and that variables that explain NPL or reserve dynamics are not always relevant to explain variation in coverage ratios.

Moreover, as we investigate the dynamics of coverage ratio components too, we are able to explore the mechanisms through which banks protect themselves against credit losses in response to shocks. This allows us, for example, to draw some conclusions on whether coverage policies are driven by accounting rather than prudential considerations and which policy measures may help foster loan loss coverage policies.

The remainder of the paper is structured as follows. Section 2 provides background details on the main measures taken to enhance loss coverage for NPLs and the reasons why it is important for banks to build up adequate coverage ratios. Section 3 illustrates the data and provides descriptive statistics for our sample. Section 4 and 5 empirically investigate the main sources of variation in coverage ratios and their components. We first focus on micro-level factors (Section 4) and then extend the analysis by using macro-level data (Section 5). Section 6 concludes.

2. NPLs and coverage ratios: Economic importance and institutional background

[5] Previous studies on LLPs discuss the role of discretion (Liu and Ryan, 2006; Bushman and Williams, 2012; Norden and Stoian, 2013; Beatty and Liao, 2014, and literature therein), as well as their timeliness and contribution to procyclical lending (Huizinga and Laeven, 2019; Laeven and Majnoni, 2003; Beatty and Liao, 2011; Nicoletti, 2018). Berger and De Young (1997), Nkusu (2011), Klein (2013), and Beck et al. (2015) among others, study the determinants of NPLs.

This section describes the supervisory initiatives introduced in recent years to enhance coverage ratios and briefly explains the role of coverage ratios as prudential tools.

2.1. Recent measures to enhance loss coverage for NPLs

NPLs have recently become a key priority for prudential authorities in Europe because of their negative effects on the stability and growth of both individual banks and the banking system as a whole. From a micro perspective, a high stock of NPLs may cast doubts on the quality of a bank's assets, thus making bank funding more expensive. This may in turn impede lending as banks with poor asset quality may seek to regain adequate capital ratios by deleveraging and cutting back on lending rather than by raising new equity. Finally, high NPL ratios can also distort bank managers' incentives in that troubled loans may increase moral hazard and favor excessive risk taking because of eroding bank capital (Bruno and Marino, 2019). From a more macro perspective, a high level of NPLs may also generate negative externalities at the system level, so that banks operating in a high NPL country may be seen in general as weaker relative to banks operating in a country with lower stocks of troubled assets (ESRB, 2019).

NPLs in European banks skyrocketed to unprecedented levels in the wake of the global financial crisis and have decreased only recently thanks in part to the pressure of the European supervisors. According to the EBA, the NPL ratio of European Union (EU) financial institutions has decreased on average from 6% as of mid-2015 to 3% as of mid-2019. Nevertheless, there are still significant discrepancies across banks and countries, with the aggregate level of NPLs in EU banks remaining very high (over 600 billion euros as of June 2019) and the gap versus international peers remaining striking, making EU banks more vulnerable than their international peers to the repercussions of poor asset quality. [6]

As argued by Constancio (2007), one of main concerns in dealing with the surge of NPLs has been the absence of common provisioning practices in Europe. This has contributed to the large variation in NPLs and coverage ratios across banks and countries, and has also impeded benchmarking and peer comparison as supervisory practice. To ensure financial stability the need to implement measures aiming to harmonize provisioning practices and enhance loss coverage have grown (Stamegna, 2019; ECB, 2019).

To strengthen the supervisory approach to NPLs, in March 2017, the ECB released guidelines on how to manage and provision for problem loans, complemented with quantitative indicators on the minimum levels of prudential provisions, based on the vintage and the degree of collateralization of the

[6] According to World Bank data, the NPL ratio was 1% in the US at the end 2018.

non-performing exposures (ECB, 2018). One year later, in July 2018, the ECB announced the decision to set bank-specific supervisory expectations for the provisioning of NPLs as part of the supervisory dialogue. The aim was to harmonize the degree of loss coverage over the medium term across comparable banks.

Along the same lines, in March 2018, the European Commission adopted a comprehensive package of measures that included a proposal to introduce common minimum coverage levels for newly originated loans that become non-performing. In April 2019, an amendment to the European capital regulatory framework, the "prudential backstop", required banks to have minimum loss coverage for non-performing exposures and to deduct from their own funds (common equity tier 1 capital) those not sufficiently covered.

To complete the picture, the accounting standard IFRS 9, introduced in January 2018, changed the impairment recognition by requiring banks, in essence, to make larger and timelier provisions based on the amount of "expected losses". Until the introduction of IFRS 9, banks in most European countries accumulated provisions according to a backward-looking approach, reflecting "incurred" credit losses (Cohen and Edwards, 2017).[7] Ideally, under the new accounting standard, provisions would better anticipate deteriorating economic conditions that may affect a borrower's ability to repay. In such a way, provisions could be used effectively to cover expected losses, instead of bank capital acting as a buffer against unexpected losses (Laeven and Majnoni, 2003). This is for example what is happening in the COVID-19 crisis, where banks have started accumulating a large amount of provisions in anticipation of future losses on their stocks of loans.

The switch to the new standard has been an important step in reconciling the perspective of accounting standard setters and bank regulators. Losses on NPLs are in fact subjected to both accounting standards and prudential regulation with different perspectives, especially before the IFRS 9 introduction. The former emphasizes transparency of financial statements, the latter emphasizes safety and soundness. From the perspective of the accounting rules, loan loss provisions have an overall detrimental effect on earnings and regulatory capital.[8] Because these are at the discretion of bank managers, there is potential for banks to provision more or less than necessary as a way to smooth their income and capital, as we will

[7] There are some exceptions. Notably, Spanish bank regulators introduced a forward–looking provisioning regime in 2000, meant to address procyclicality issues, which led to more timely and higher general provisions (de Lis et al., 2001; Jiménez et al., 2017).

[8] The actual effect on bank capital of provisioning is hard to determine, because the regulatory implications of provisions varies according to the approach used by banks for calculating capital requirements, and on the nature of bank provisions (namely, general vs. specific provisions). See Bruno and Carletti (2017) for a concise discussion on the effects of provisions on bank capital.

discuss in Section 4.1. On one hand this would introduce discretionary modifications to earnings and reduce comparability across firms (Walter, 1991). On the other hand, from a prudential perspective, higher provisioning may reflect a more cautious approach to building up large reserves prior to future losses.

2.2. Coverage ratio as a prudential tool

The initiatives illustrated above show that coverage ratios have gained relevance as a key prudential and monitoring tool to shield banks' balance sheets. Why is it desirable for regulatory and supervisory purposes to promote high loan loss coverage? The answer is that adequate coverage ratios can help banks mitigate most of the concerns associated with high NPLs.

Adequate loan loss reserves, and thus high coverage ratios, for a given level of NPLs, enhance banks' safety and soundness by protecting bank capital when losses materialize (Wheeler, 2019). Specifically, loan loss reserves are a "contra-asset" account, which reduces the loans by the amount the bank expects to lose when some portion of the loans are not repaid. Periodically, the bank managers decide how much to add to the LLR account, and record this amount as an expense item on the profit and loss account through "provisions for loan losses". This allows banks to recognize the estimated loss even before the actual loss can be determined with accuracy and certainty. To the extent that credit risk is not under-estimated and allowances are adequate to cover for the actual loss, by building adequate coverage ratios banks protect their capital and preserve their capacity to provide credit to the economy (Beatty and Liao, 2011).[9]

High coverage ratios also help to make banks' balance sheet more transparent. In the traditional banking literature (e.g., Diamond and Dybvig, 1983), loans are illiquid and untraded contracts generating cash flows that are hard to predict. In the absence of a true market price, the loan fair value is approximated through the process of provisioning. The process of accumulating provisions is, in fact, equivalent to reducing the face value of the loan to its present value, taking into account the allowance built up over time (Song, 2002). If loan loss allowances were underestimated, bank assets and capital ratios would be overvalued and balance sheets would be distorted.

[9] The NPL Guidance also stresses the importance of timely provisioning related to NPLs, as "these serve to strengthen banks' balance sheets, enabling them to (re)focus on their core business, most notably lending to the economy" (ECB, 2018).

Relatedly, because high loan loss coverage corresponds, *de facto*, to low loan net book value, it follows that reporting high coverage ratios is also a precondition to make the asset disposal more likely and reduce the bid-ask spread between sellers and buyers (Fell et al., 2016). However, anecdotal evidence and market practices show that, on average, coverage ratios in European banks are still inadequate if compared to actual recovery rates or haircuts applied as an effect of NPL resolution.[10] This points to the importance of increasing coverage ratios in order to reduce the negative impact of credit losses on capital.

In sum, coverage ratios are important tools to ensure the safety and soundness of the banking sector, enhance the transparency of banks' balance sheets and favor the disposal of NPLs. Yet, as we will show below, they show important variation both across banks and countries. Because of this, a number of policy measures have been introduced in recent years aiming at increasing the level of coverage ratios and decreasing their dispersion. In what follow we analyze the determinants of coverage ratios in Europe, as well as of their components, and derive implications as to which policies may be more effective.

3. Data and summary statistics

We collect annual bank-level data from the S&P Global Market Intelligence Platform (S&P Global). The dataset spans the years 2010–2017 and covers all EU countries as of 2017. Following Eber and Minoiu (2016), we collect data at the highest consolidation level. To avoid including small banks that could introduce noise, we only keep banks that are being classified as medium-sized and large according to the ECB definition.[11] Given the purpose of the analysis, we also drop the institutions whose commercial banking business is negligible from the sample.[12] All variables are winsorized at 2.5% and 97.5%. The final

[10] In the context of the NAMA, the asset management company established in Ireland in 2009, assets were priced with a 57% haircut, with an average haircut on loan portfolios ranging from 43% to 61%. In the case of SAREB, the Spanish asset management company established in 2012, total assets were valued with a 53% haircut, with large discrepancy by loan type (Medina Cas and Peresa, 2016). Looking at Italy, the recovery rate on NPLs is estimated between 41% (Carpinelli et al., 2016) and 47% (Ciavoliello et al., 2016), indicating an average haircut of about 60%.

[11] The ECB labels as large those institutions with assets greater than 0.5% of total consolidated assets of European Union banks and medium–sized as those with assets between 0.5% and 0.005%.

[12] We delete institutions with a loan-to-asset ratio and a deposit-to-asset ratio smaller than 20%, those not classified as 'bank' or 'savings bank/thrift/mutual', as well as those that, although being classified as banks by S&P Global, may operate not in a pure commercial manner because for example of ownership (e.g., government-owned banks) or scope (e.g., asset management companies).

sample contains 441 banks, representing around 70% of banking assets in Europe. Table A.1 reports the breakdown of observations and banks in our sample.[13]

Figures 1 to 3 explore trends in NPLs, LLRs, and coverage ratios in on our sample. Figure 1 shows that the evolution of the average coverage ratio over all countries and in high-NPL countries (low-NPL countries), defined as those with NPL/TA above (below) the sample mean.[14] In both groups of countries, coverage ratios have trended up since the sovereign debt crisis in 2010–2012 and, again, after the introduction of the single supervisory mechanism (SSM) in 2014. Overall, European banks have progressively increased their coverage ratios, partly as a managerial response to asset quality deterioration and partly due to stricter supervisory and market scrutiny.[15]

Throughout our sample period, high-NPL countries tend to report coverage ratios below the sample average, although the gap has progressively narrowed over time. In fact, most of the time variation in coverage ratios seems to be explained by high-NPL countries, as they have increased from nearly 35% to 55% in 2010-2017, as opposed to low-NPL countries whose average coverage ratio moved from 45% to 55%. Figures 2 and 3 show the dynamics of the components of the coverage ratio for high and low-NPL countries, respectively. By comparing Figure 1 with Figures 2 and 3, it emerges that while the dynamics of LLRs and NPLs are similar, they are different from those of coverage ratios.

Figures 4 and 5 confirm the presence of large cross-sectional variability in asset quality and coverage ratios, respectively, both across countries and within the same country (see also Table A.1 for a sample composition in terms of per–country average coverage ratios and their components). Figure 4 shows that countries with higher median NPLs also have a larger dispersion in NPL/TA across banks. By comparing the two figures, no obvious country-level mapping emerges between the quality of bank loans

[13] As it emerges from Table A.1, German banks are over-represented in terms of number of institutions in our sample. This is common in the empirical literature on European banks (see Altavilla et al., 2017, among others) and reflects the highly fragmented nature of the German banking system. To check whether this has implications, we have re-run the analysis on a sample excluding German banks. Results, available upon request, remain robust.

[14] Our definition of high-NPL countries is time-varying, with some countries coming in only for part of the sample. All countries in which the NPL ratio exceeds 10% in 2016 (in accordance with the definition of the ESRB, 2017) are consistently covered. These countries are the following, in order of descending NPL ratio: Greece, Cyprus, Portugal, Italy, Slovenia, Ireland, Bulgaria, Hungary, Romania, and Croatia.

[15] This may be due to stricter supervisory and regulatory scrutiny in relation to the ECB's asset quality exercises, increased market pressure, as well as a deterioration of collateral values (Council of the European Commission, 2017).

and the level of coverage. This suggests that although differences in asset quality may contribute to explain heterogeneity in European banks' coverage ratios, other factors may also play a role.[16]

Descriptive statistics and correlations for all the variables are shown in Table 2 and Table A.2, respectively. The average bank in our sample is a traditional commercial bank, whose core business is lending (the average loan to asset ratio is 65%) and whose main source of funds are customer deposits (the deposits to assets ratio averages 66%). As far as bank asset quality is concerned, the NPL to total asset ratio averages at about 4%. The average coverage ratio is 51%, with large variation across banks (the minimum coverage ratio being 10% and the maximum 89%). These numbers are comparable to those reported in aggregate statistics (ECB, 2016; EBA, 2018).

Looking at measures of bank capitalization, the CET 1 regulatory capital ratio is on average 15%, well above the Basel III minimum requirement of 8.5% including the capital conservation buffer. The average ROAA is around zero, confirming that low profitability has been a major source of concerns for European banks and that high NPLs have been an important cause of low profitability in European banks (Altavilla et al., 2018).

Table 2 also shows descriptive statistics for the set of macro variables we consider, namely institutional variables, including the depth of the NPL secondary market, and business/financial cycle indicators. The former include two indices to account for the regulatory and judicial environment, namely the Regulatory Quality index and the Rule of Law index, both published by the World Bank, and the a series of macroprudential variables, grouped in a Macroprudential index as in Cerutti et al. (2017) macroprudential policy dataset. The latter include business cycle indicators such as real GDP growth and unemployment rate, variables related to the financial cycle, such as asset price growth (i.e., house and stock prices), and private credit to GDP ratio, as well as the short term interest rate. A description of these macro variables, together with the relative hypotheses, is given in Section 5.

4. Exploiting the cross section of banks: micro-level analysis

In this section we analyze the role of the micro bank-specific variables in explaining coverage ratios. We start with illustrating the main specification and testable predictions, and then present the results.

[16] An EBA report on NPLs also shows that the correlation between these assets and coverage ratios is low over time, with a correlation coefficient close to 0 at least since September 2014 (EBA, 2016).

4.1 Baseline specification, main variables, and testable predictions

To explore the link between coverage ratios and bank specific characteristics we first exploit our sample heterogeneity at the micro-level. Looking simultaneously at the coverage ratio and its components, loan loss reserves and non-performing loans (both scaled by total assets), enables us to better understand the mechanisms by which banks set coverage ratios, over and above the accounting identification of impaired loans. Our key dependent variable is the coverage ratio, in addition, we also use its components as additional dependent variables in separate models.[17]

We estimate the following regression having LLRs, NPLs and coverage ratios as dependent variables in separate models:

$$Y_{i,k,t} = \mu_i + \gamma_{k,t} + \beta X_{i,k,t-1} + \varepsilon_{i,k,t}, \qquad (1)$$

where $i = 1, \ldots, N$, $k = 1, \ldots, K$ and $t = 1, \ldots, T$, with i being the bank, k being the country, and t being the year. $Y_{i,k,t}$ is our dependent variable, which can be coverage ratio or its components, that is loan loss reserves or NPLs over total assets. The vector $X_{i,k,t-1}$ includes bank-level variables to account for bank specific factors that can be relevant in determining the coverage ratio and its components. The equation includes bank and country-year fixed effects (μ_i and $\gamma_{k,t}$, respectively).[18] In one specification, we replace bank fixed effects with various time-invariant characteristics, as we explain further below and later in Section 4.2. All explanatory variables (with the exception of the change in NPLs and loan growth) are lagged by one year to mitigate concerns about reverse causality. When $Y_{i,k,t}$ equals the ratio of LLRs to total assets (NPLs to total assets), we remove the lagged LLRs to total assets (NPLs to total assets) as explanatory variable.

In identifying the bank-specific drivers of banks' coverage policy, we draw primarily on the literature which examines the determinants of provisioning and NPLs. We group our independent variables in four main categories: credit risk, funding, bank performance, and forward looking.

[17] We are aware that across jurisdictions and banks there may be different definition of NPLs (Baudino et al., 2018). A harmonized definition of NPLs was however introduced in 2014 by the EBA, by which non-performing loans are those that satisfy either of the following criteria: (a) exposures that are more than 90 days past due; and (b) the debtor is assessed as unlikely to pay its credit obligations in full without realisation of collateral. Unfortunately, the breakdown of the NPL aggregate is unavailable for most banks in our sample.

[18] The inclusion of bank and country fixed effects is also important to absorb the variation in coverage ratios due to possibly different definitions of NPLs across banks and jurisdictions.

We start with a large set of credit-risk related variables. In the literature on bank provisioning these factors are referred to as non-discretionary, as opposed to (discretionary) characteristics accounting for different management objectives (see Beatty and Liao, 2014, among others). Specifically, we include measures of asset quality such as the level of loan loss reserves as well as the level and the change of NPLs (scaled by total assets). *Ceteris paribus*, we expect poorer asset quality to be associated with higher loan loss reserves, as banks with higher NPLs should be more prone to increase loss coverage for the reasons discussed in Section 2. In one specification, in the spirit of Bushman and Williams (2012), we also test whether banks' coverage policy includes forward-looking considerations, which we model by including next year's change in non-performing loans, to account for (potential) future losses. We then include variables measuring the relevance of the lending business (the share of gross loans over total assets) as well as the growth of gross loan as other potential factors affecting credit risk and therefore banks' loss coverage policies (Bouvatier and Lepetit, 2012; Nicoletti, 2018). The idea is that banks that are more willing to invest their funds in loans (rather than, e.g., securities) are more exposed to credit risk (Keeton and Morris, 1987). Also, excessive credit growth may be associated with more risky lending, and hence with higher NPLs in the future (Jiménez and Saurina, 2006; Huizinga and Laeven, 2019). It follows that a larger share of loans to total assets and higher credit growth should favor a more prudent coverage policy and therefore higher coverage ratios. Finally, we control for size, measured by the natural logarithm of total assets, as aggregate statistics show that smaller banks tend to report higher coverage ratios (EBA, 2018). More generally, prior research has shown that size is a relevant determinant of lending and risk taking (see Kishan and Opiela, 2000, among others), and, thus, it may also explain banks' coverage ratios and their components.

To investigate the role played by bank funding structure, we include measures of capitalization, by using the common equity tier 1 (CET 1) capital ratio, and reliance on deposits, proxied by the share of customer deposits to total assets. Capital plays contrasting roles in terms of coverage ratios. Previous studies argue that bank managers may exploit discretion in provisioning not only to smooth income, but also to manage capital (see, among others, Liu and Ryan, 2006 and Beatty and Liao, 2014, and literature therein). It follows that capital-constrained banks may have an incentive to use provisions to achieve regulatory capital targets (Andries et al., 2017). This occurs because provisions have a mechanical negative effect on banks' capital, by reducing earnings.

These arguments point to a positive relationship between capitalization and provisioning, as weak banks would have the incentive to hold back on LLPs and under-reserve in order to preserve regulatory capital. In addition, according to the "moral hazard" hypothesis (Keeton and Morris, 1987), undercapitalized banks are more prone to gamble for resurrection and thus increase the riskiness of their loan portfolio compared to stronger banks, also by lending to zombie firms (Schivardi et al., 2018, and literature therein). Taken together, these theories imply a positive correlation between capital and coverage ratios, through both the effects on reserves and NPL levels.

An alternative view would instead justify the existence of a negative nexus between coverage ratios and regulatory capital as the two balance-sheet items are seen as substitutable buffers against potential losses. In this view, low capitalized banks may have the incentive to increase loan loss coverage to partly compensate for their lack of capital (Norden and Stoian, 2013). Or, to change perspective, better capitalized banks would be in a more comfortable position to absorb shocks prompted by the deterioration of the loan portfolio. As such, these banks would have less incentives to set high coverage ratios.

The relevance of deposits may also help explain banks' reserving practices. In line with Calomiris and Kahn (1991), we expect that banks with a larger share of demandable debt, being more exposed to market discipline, have stronger incentives to report high coverage ratios compared to banks that rely less on deposits.[19]

We then test whether bank performance, as measured in terms of profitability (proxied by the return on average assets, ROAA) and efficiency (proxied by the cost-to-income ratio, i.e., the ratio of operating expenses over operating income) influences coverage ratios. According to the income-smoothing hypothesis (see Liu and Ryan, 2006 and Beatty and Liao, 2014, and literature therein), when earnings are low, provisions are deliberately understated to mitigate the adverse effect of other factors on earnings, in contrast to situations when earnings are high. Conversely, banks can smooth their earnings by drawing from loan loss reserves if actual losses exceed expected losses.[20] This results in a systematic under (over)-reserving in banks with low(high) profits. We therefore expect a positive correlation between ROAA and coverage ratios.

[19] A positive association between the deposit to asset ratio and coverage ratio is also in line with Drechsler et al. (2018). They argue that deposits effectively behave as term liabilities because banks are able to exert market power. They thus optimally invest into (risky) long-term assets. Hence, any positive correlation between deposits and coverage ratios could reflect some bank assets' characteristic not directly captured by our variables.

[20] As bank profitability and GDP growth tend to be positively related, income smoothing would be implicitly forward-looking in nature and can mitigate pro-cyclicality (Laeven and Majnoni, 2003; Bushman and Williams, 2012).

As for cost efficiency, in the literature on NPL determinants a high cost-to-income ratio can be associated with either higher or lower troublesome loans, according to whether the "bad management" prevail over the "skimping" hypothesis (Berger and De Young, 1997). Under the bad management hypothesis, low cost efficiency (i.e., high cost-to-income ratios) is a signal of poor management practices, thus implying lower portfolio quality as a result of poor screening and monitoring. On the contrary, under the skimping hypothesis, high cost-to-income ratios are associated with lower NPLs, as more resources are allocated to the monitoring of credit risk. As a result, when the cost-to-income increases, we then expect higher NPLs and, *ceteris paribus*, lower coverage ratios if the bad management view prevails, as opposed to when the skimping hypothesis dominates.

Another strand of literature (Ristolainen, 2018) links more directly the effect of bank performance on coverage ratios through banks' incentives to under-report NPLs or to under-reserve, which would be stronger in less profitable and less efficient banks. Consistent with this view, we expect lower coverage ratios when bank performance worsens.

Finally, we include a number of time-invariant bank characteristics (in the form of dummies) when removing the bank fixed effects in one specification. These variables include: *Significant*, to account for the institutions included in the 2014 Comprehensive Assessment exercise; *Listed* and *Saving, Mutual or Thrift,* to account for differences across bank owners/business type; *International Financial Reporting Standards (IFRS)*, to control for possible heterogeneity in reporting practices. In addition, we include a set of dummies that capture structural aspects related to banks' loan loss reserve policy, asset quality and lending strategy, size, funding, and performance identifying banks that rank in the top decile of the distribution of the following variables: LLR/TA, NPL/TA, Gross Loans/TA, Log(TA), Deposits/TA, CET1 ratio, ROAA and cost-to-income ratio.[21] Based on these reference variables, we classify banks as *Well reserved*, *Risky*, *Loan-based*, *Large*, *Deposit-based*, *Sound*, *Profitable* and *Inefficient*.

4.2 Results

From a policy maker's view point it is important to understand which factors explain most of the variation in loan loss coverage policy. To gauge these factors, we proceed in steps.

4.2.1 Micro time-varying and invariant variables

[21] The dummies are time invariant since they are constructed based on average values for the entire length of the sample.

As a preliminary analysis, we run our main regression on the coverage ratio by including only fixed effects at the bank and the country-year level. As shown in Table 3, the regression including only bank fixed effects has an adjusted r-squared of 0.8, while the one with bank and country-year fixed effects has an adjusted r-squared of 0.82. These results show that most of the variation of the coverage ratio is explained by time-invariant bank characteristics and that the additional fixed effects only mildly improve the statistical fit. In terms of policy implications, it follows that bank characteristics matter more than country specificities in explaining bank loan loss coverage policies, and that therefore policy makers concerned about coverage ratios should first and foremost strengthen microprudential oversight.

We then analyze which of the (time-varying and time-invariant) bank characteristics help explain variations in the coverage ratio and its components. Columns 1 to 3 of Table 4 present the results for the baseline investigation on the main micro drivers of NPLs, LLRs and coverage ratios, respectively, where bank fixed effects are replaced by the time-invariant characteristics described in Section 4.1.

We find that among the structural components, significant banks tend to report lower coverage ratios, as also found in Ristolainen (2018), possibly because of too-big-too fail motives. At the same time, listed banks show significantly higher coverage ratios, perhaps as an effect of closer investor scrutiny for these banks than for unlisted banks.

Turning to the dummy variables used to identify the time-invariant component of our main baseline variables, we find that well reserved and risky banks report lower and higher coverage ratios than the average bank, respectively. This evidence suggests that considering loan loss reserves and NPLs separately can be misleading, supporting the argument that the NPL stock should be looked at only in conjunction with the associated degree of coverage (Constâncio, 2017). We also find that loan-based and sound (well capitalized) banks tend to have lower coverage ratios. The latter result points to a substitution effect between capitalization and loan loss coverage for banks with high capital levels, as suggested in Norden and Stoian (2013).

Interestingly, Table 4 also shows that the large set of bank characteristics included in the analysis explains the variation of NPLs and LLRs well (the adjusted r-squared in Columns 1 and 2 is above 0.9), but it seems to be less powerful in explaining the variation in the coverage ratio (the adjusted r-squared in Column 3 is 0.56). This finding indicates again that looking at only the dynamics of loan loss reserving and NPLs is not sufficient to fully understand the dynamics of coverage ratios. It also suggests that there may be omitted variables which explain the way banks set their coverage ratio. These variables plausibly pertain to the individual bank's managerial sphere and are, therefore, unobservable (from a modeler's point of view) or are hard to identify.

As a next step we include bank fixed effects to account for bank-specific time invariant characteristics, including unobservable ones. In Table 4 Columns 4 to 6 present the results for our baseline specification, results are broadly consistent with those without bank fixed effects. Among the time-varying variables, credit risk variables are important to explain coverage policy. We find in particular that the relationship between the level and the change of NPLs and coverage ratio is negative (Column 6), while, as in Huizinga and Laeven (2019), there is a strong positive relationship between asset quality and LLRs (Column 5). This means that although banks tend to react to higher NPLs by increasing loan loss reserves, such an increase does not seem adequate to compensate for the larger amount of NPLs. As a result, when the loan portfolio quality deteriorates, coverage ratios reduce.

We find that higher credit growth is associated with larger loan loss reserves and higher coverage ratios, despite the negative relationship between credit expansion and NPLs. This last result suggests that, in line with Jiménez and Saurina (2006) and Huizinga and Laeven (2019), when the loan portfolio expands, banks prudently enhance their loan loss coverage by anticipating higher (potential) future losses, independent of the impact higher credit growth has on the NPL/TA ratio in the short run.

Among the variables capturing bank funding structure, capital is positively related to coverage ratios, although only at the 10% level, but not with the individual components. This suggest that capital and coverage ratios are not substitute approaches to deal with loan losses, except perhaps for banks with very high capital as shown in Column 3 of the table. Concerning bank performance, profitability explains only the dynamics of the individual components but not coverage ratios directly, while the degree of efficiency, as captured by the level of the cost-to-income ratio, is negatively correlated with both NPLs and coverage ratios. Overall, these results provide some support to the view that lower performance increases banks' incentives to under-report NPLs and to under-reserve, as found in Ristolainen (2018).

As robustness check (see Table A.3), we replace our asset quality indicator with the NPLs to total loans ratio, the ROAA with the return on average equity (ROAE), the CET1 ratio with the Tier 1 ratio. Results remain consistent with the baseline specification.

4.2.2 Forward-looking variables and high-NPL banks

Next, we extend our baseline specification to account for the forward-looking behavior of banks and investigate the behavior of high-NPL banks. Results are shown in Table 5. Columns 1 and 2 report results from a specification where we add the change in NPLs at t + 1, to account for (potential) future losses, to the baseline. We find a strong positive association between this forward-looking measure of asset quality and coverage ratios.[22] This finding reinforces the interpretation of our results on credit growth, suggesting that coverage ratios work more as a prudential (forward-looking) buffer than merely (and backward-looking) a booking account.

Columns 3 and 4 in Table 5 explore the differential behavior of banks with the highest levels of NPLs. On one hand we expect that banks with high NPLs should face higher expected losses and should therefore be more in need of setting up higher coverage ratios to protect their balance sheets. On the other hand, because provisions to loan-loss reserves would further reduce earnings and capital, high-NPL banks may have more incentives to under-provision for potential losses when asset quality further deteriorates, or when profits and capital decrease relative to banks with lower NPL ratios (Ristolainen, 2018). To exploit the large discrepancies among NPLs ratios we focus on banks in the top decile of the NPL/TA ratio distribution by including *High NPL* dummy and its interaction with the share of NPLs to total assets, CET1 ratio, and ROAA. Note that this *High NPL* dummy variable is now time-varying, in contrast with the dummy variable *Risky* used before representing banks with structurally high NPLs levels during the whole sample.

Results in Columns 3 and 4 show that while higher NPLs are in general associated with reduced coverage ratios, in high-NPL banks this correlation is significantly less negative, pointing to a non-linear relationship between asset quality and coverage ratios. While banks are generally unable (or unwilling) to adjust their loan-losses at the same pace as asset quality deteriorates, banks facing a very high level of credit risk try to restore an adequate level of coverage. This finding may be driven by particularly strong supervisory pressure or peer effects. The result confirms the one found for banks with structurally high levels of NPLs in Table 4.

Turning to capitalization we uncover a positive association between the level of capital and loan loss reserves in high-NPL banks, but with no differential effect on coverage ratios. As for the nexus between profitability and coverage ratio, we find a significant and negative correlation, suggesting that high-NPL banks tend relatively more to use their profits in other ways than to increase reserves and coverage ratios, consistent with a pro-cyclical behavior of bank provisioning (Huizinga and Laeven, 2019).

[22] In untabulated results, available upon request, we replace the change in NPLs at t+1 with the lead of the NPL to total asset ratio. The positive effect on coverage ratios is confirmed.

5. Exploring macro-level data

In this section we exploit the richness of country characteristics to better explain the variation in coverage ratios across countries. We replace the country-year fixed effects with a large set of time varying macro variables related to institutional/governance rules and macroprudential policy to analyze their role as potential drivers of banks' coverage choices. In doing this, we also consider separately the specificities of high-NPL countries and the role of a secondary market where NPLs can be sold.

5.1. Specification and variables

We estimate the following regression having LLRs, NPLs and coverage ratios as dependent variables in separate models:

$$Y_{i,k,t} = \mu_i + \lambda_t + \beta_1 X_{i,k,t-1} + \beta_2 Z_{k,t-1} + \varepsilon_{i,k,t}, \qquad (2)$$

where $X_{i,k,t-1}$ includes lagged bank-level variables as illustrated in Section 4 and $Z_{k,t-1}$ comprises the lagged time-varying macro-level factors capturing three dimensions: regulatory quality, rule of law, and macroprudential stringency. Table 2 reports aggregate statistics for all the macrovariables included in the analysis. We saturate the specification with bank and year fixed effects (μ_i and λ_t, respectively).

In the spirit of Andries et al. (2017), we include *Regulatory Quality* as a measure of the government's ability to formulate and implement policies and regulations. To capture the quality of the judicial system, we include an index of *Rule of Law* capturing agents' confidence in rules, quality of contract enforcement, property rights and courts. Both variables are published by the World Bank, based on an annual survey. We expect better regulatory quality as represented by higher values of *Regulatory Quality* to be associated with more prudent coverage policy and thus higher coverage ratios. We also expect more stringent (higher) *Rule of Law* to be associated with lower coverage needs, as for example banks may recover NPLs more quickly and efficiently when the legal and judicial framework is strengthened.

To analyze the role played by macroprudential policy, we include the 2018 update of the country-specific prudential measures as derived from the Cerutti et al. (2017) macroprudential policy dataset. We start with the broadest index available in the dataset, the so-called *Macroprudential Index*. This covers three borrower-targeted and nine financial-institution-targeted instruments, therefore taking on values between 0 and 12, where 0 means that none of the instruments are in place and 12 means that all of them are in place. Hence, the higher the index, the more stringent the implementation of macroprudential measures in the respective country. We then replace the index by some of its subcomponents. Based on anecdotal evidence in Walter (1991) and prior research on the effects of macro factors on banks provisioning (Jiménez et al., 2017 and Andries et al., 2017, among others), we focus on those ones that are more likely to affect banks' coverage ratios, namely: *Dynamic loan-loss provisioning* as a measure of provisioning policies, *Capital Surcharges on Systemically Important Financial Institutions (SIFI)* as a measure of capital buffers, *Levy/Tax on Financial Institutions (FI)*, and *Loan-to-Value (LTV) Ratio Caps* capturing the limits to borrowing.

We also include a *High-NPL Country* dummy, to account for banks from countries with an above sample average level of NPLs. [23] All things being equal, banks from countries affected by high levels of NPLs may behave differently from the average sample bank. Most of these countries have in fact weaker institutional frameworks and as such banks may face more impediments in resolving NPLs (Aiyar et al., 2015; ECB, 2016). This may delay NPLs disposals and induce distortions in banks' provision policies.

Over the last years, high-NPL countries have been under particularly close scrutiny from national and supranational authorities, and banks from these countries have been required to undertake specific efforts to strengthen their balance sheets. It follows that we expect any regulatory intervention in these countries to lead to a relatively stronger reaction by banks located in these countries. [24] To investigate whether this is the case, we interact the high-NPL country dummy with all our proxies for country governance and policy.

Finally, we also control for the business and financial cycle by including a broad range of macroeconomic and financial variables derived from the literature on NPL determinants (Nkusu, 2011; Klein, 2013; Beck et al., 2015) and provisioning procyclicality (Laeven and Majnoni, 2003; Beatty and Liao, 2014). In particular, *Real GDP growth* and the *Unemployment rate* are used as indicators of general macroeconomic performance. *House Price change* and *Stock Price change* help explain differences in asset quality, e.g. via wealth effects among borrowers or via a decreased value of collateral. *Private Sector Credit-to-GDP* captures the aggregate debt burden of households and businesses. Finally we control for *Short term interest rates* as monetary policy may also influence asset quality and loan loss coverage policy.

[23] The definition of high-NPL country is the one introduced in Section 3, i.e. a time-varying definition (see footnote 13 for details).

[24] In fact, the policies and practices in jurisdictions not afflicted by high NPLs "are not expected to be as prescriptive or coordinated as those in jurisdictions currently reacting to high levels of NPLs" ECB (2016).

5.2. Results

Table 6 shows the results of our investigation on the role that quality and stringency of the institutional and regulatory framework play on banks' coverage policy. For sake of space, all the bank-specific variables and the set of macro variables which capture the economic and financial cycle are included in the analysis, but not explicitly reported in the table.

Among all the macro variables considered, only the macroprudential index is positively associated with both reserves and coverage ratios (Columns 2 and 3). Among the components of this index, dynamic loan-loss provisioning is associated with lower NPLs and higher coverage ratios (see Columns 4 and 6). This indicates that when measures to address pro-cyclical provisioning are in place, banks are better able to increase coverage ratios. We also find evidence that taxation on financial institutions is associated with higher coverage ratios (Column 6), plausibly because of the possibility of higher deductions associated with larger provisions (Andries et al., 2017).[25]

Interestingly, in countries most affected by NPL issues, stricter rule of law is associated with lower NPLs, indicating that better quality enforcement or more efficient courts are relatively more beneficial for NPL accumulation presumably as they entail a quicker recovery phase (Columns 1 and 4). In line with this, stricter rule of law is also related to lower coverage ratio (Columns 3 and 6), perhaps because of lower reserve needs when recoveries are higher.

Among the various macroprudential measures, capital surcharges for systemically important institutions have the strongest impact in high-NPL countries and are associated with lower NPLs and coverage ratios (Columns 4 and 6). This finding is in line with previous research on stricter capital regulation which finds that when banks comply with stricter capital rules deleveraging and de-risking strategies are more likely (Gropp et al., 2019). This mechanism is likely to hold in high-NPL countries where banks presumably have a higher incentive to retain earnings to comply with the new rules rather than to increase provisioning.

As a final comment, it is important to note that although the bank-specific variables are not included in Table 7 for sake of space, they remain the most important determinants of coverage ratios. This is evident in Table A.4 where we carry out a Shapley decomposition to analyze variance explained by the micro and macro determinants we use in our regressions.

[25] Although at different rates, the majority of EA countries "acknowledge tax deductions for LLPs, write-offs and collateral sales". (ECB, 2016 and 2017b).

5.3 Extension: NPL secondary market and coverage policy

One of the responses most often cited by banks as an impediment to the NPL resolution is the lack of a market to sell NPLs (EBA, 2019). Although relatively underdeveloped in relation to the high NPL stock in some jurisdictions in Europe, NPLs transactions have progressively increased over the last years, varying from 11 billion euros in 2010 to nearly 100 billion euros as of end 2017, according to PwC reports. Transactions are concentrated in a few countries, i.e., Ireland, Germany, Spain, and UK, and more recently, Italy (the largest market place since 2016). [26] Figure 6 shows the value of NPL transactions by country in 2010–2017.

The market for distressed assets is clearly a market for lemons à la Akerlof, being characterized by high information asymmetries and large bid-ask spreads between sellers and buyers (Fell et al., 2016). High coverage ratios can help make the disposal of loans more likely by reducing the bid-ask spread and the loss a bank takes as a consequence of the NPL sale (see also the discussion in Section 2). We therefore expect deeper markets to be associated with higher coverage ratios as a pre-condition to access the market (see also the discussion in Section 2).

To test this hypothesis, in Table 7, we expand our micro-macro baseline regression to account for the relevance of the NPL secondary market in a given country. We first include the variable *NPL Secondary Market Transactions / TA* to measure the share of NPL transactions over the total banking assets at the country level to proxy the degree of development of the market (Columns 1 to 3). Because the volume of trades is concentrated only in some countries, we also include two categorical variables to account for *Medium* and *Large NPL Secondary Market*, by splitting the sample into terciles (based on the share of NPL transactions over the total banking assets at country level). We use the lowest tercile as the reference category and test whether the other categories are associated with higher coverage ratios. We find that while LLRs are higher when transactions increase and, more generally, in medium sized and large marketplace (Columns 2 and 5), coverage ratios are significantly higher only in countries where the NPL secondary market is large (Column 6).

As a next step, we interact our measures of medium and large NPL secondary markets with the high-NPL country dummy. In line with official statistics, we find that banks from high-NPL countries report lower coverage ratios on average. We find, however, relatively larger reserves and higher coverage ratios

[26] The dataset also includes transactions for Portugal (2011), France (2012), Belgium (2013), and Netherlands (2013, 2014, 2015 and 2016) but for more limited amounts.

in banks from high-NPL countries that are featured by very active marketplaces (Columns 8 and 9). This is not surprising, as banks from high-NPL countries, are more affected by information asymmetries (see Fell et al. 2016) and therefore may need to set higher coverage ratios to access the market.

6. Conclusions

This paper explores micro and macro determinants of coverage ratio, an indicator of bank balance sheet strength that has gained increasing importance in Europe in the last few years.

Our analysis reveals some interesting findings. Bank-specific factors, and among them credit risk (including forward-looking) variables, explain most of the variation in coverage ratios. A deterioration in asset quality is associated with higher coverage ratios, but the relation is not linear, becoming less negative when banks hold very large stock of troubled assets. Overall, capitalization and coverage ratio appear to be complementary (rather than substitute) tools, where one reinforces the other.

More stringent macroprudential policy is also associated with higher coverage ratios, and interventions on time-varying/dynamic loan-loss provisioning are generally the most effective tools to increase coverage ratios. Structural factors such as the degree of development of NPL secondary markets also explain coverage ratio variation, where larger markets are associated with higher coverage ratios.

High-NPL banks as well as banks from high-NPL countries behave differently from banks less affected by credit risk issues. Coverage policies in banks from more risky countries are especially sensitive to changes in the rule of law, capital rules, and development of the NPL secondary market.

Our results are relevant for the current debate on NPLs and coverage policies. We uncover that variables that are traditionally important in explaining NPLs dynamics are not equally useful to explain variation in loan loss coverage. Bank-specific factors explain most of the variation in banks' coverage ratios, implying that microprudential supervision would be more effective in steering banks' loan loss coverage than macro policies. In terms of macro policies, some specific macroprudential levers, as well as developing loan secondary markets, seem to be effective in shaping banks' coverage. Because of the large discrepancies in asset quality across banks and countries, specific actions for high-NPL banks and high-NPL countries are recommended.

References

Aiyar, S., Bergthaler, W., Garrido, J. M., Ilyina, A., Jobst, A., Kang, K. H., Kovtun, D., Liu, Y., Monaghan, D. and Moretti, M. (2015). A strategy for resolving Europe's problem loans. IMF Staff Discussion Note, SDN/15/19, September.

Altavilla, C., Boucinha, M., and Peydró, J.L. (2018). Monetary policy and bank profitability in a low interest rate environment, Economic Policy, October: 531–586.

Andries K., Gallemore, J. and Jacob, M. (2017). The effect of corporate taxation on bank transparency: Evidence from loan loss provisions. Journal of Accounting and Economics, 63: 307–328.

Ari, A, Chen, S. and Ratnovski, L. (2020).The dynamics of non-performing loans during banking crises: a new database. ECB Working Paper 2395.

Baudino, P, Orlandi J. and Zamil, R. (2018). The identification and measurement of non-performing assets: a cross-country comparison. FSI Insights on Policy Implementation No 7, Bank for International Settlements.

Beck, R., Jakubik, P. and Piloiu, A. (2015). Key determinants of non-performing loans: new evidence from a global sample, Open Economies Review, Vol. 26(3): pp. 525–550.

Berger, A. and R. DeYoung. (1997). Problem Loans and Cost Efficiency in Commercial Banks. Journal of Banking and Finance (21): 849–870.

Beatty, A. and Liao, S. (2011). Do delays in expected loss recognition affect banks' willingness to lend? Journal of Accounting and Economics, 52 (1): 1–20.

Beatty, A. and Liao, S. (2014). Financial accounting in the banking industry: A review of the empirical literature. Journal of Accounting and Economics, 58: 339–383.

Bouvatier, V. and Lepetit, L. (2012). Provisioning rules and bank lending: A theoretical model. Journal of Financial Stability, 8(1): 25–31.

Bruno, B. and Carletti, E. (2017). Provisioning policies for non-performing loans. How to best ensure a clean balance sheet? European Parliament, In-Depth Analysis requested by the ECON committee, October.

Bruno, B. and Marino, I. (2019). How Banks Respond to NPLs. Evidence from the Euro Area. CSEF Working Papers 513, Centre for Studies in Economics and Finance (CSEF), University of Naples.

Bushman, R.M. and Williams, C.D. (2012). Accounting discretion, loan loss provisions, and discipline of banks' risk taking. Journal of Accounting and Economics, 54: 1–18.

Calomiris, C. and Kahn, C. (1991). The Role of Demandable Debt in Structuring Optimal Banking Arrangements. American Economic Review, 81(3): 497-513.

Carpinelli, L., Cascarino, G., Giacomelli, S. and Vacca, V. (2016). The Management of Non-performing Loans : A Survey Among the Main Italian Banks. Questioni di Economia e Finanza, Occasional Papers 311, Bank of Italy, 1-32.

Ciavoliello, L. G., Ciocchetta, F., Conti, F. M., Guida, I.,Rendina, A. and Santini, G. (2016). What's the Value of NPLs? Notes on Financial Stability and Supervision 3, Bank of Italy, 1-8.

Cerutti, Eugenio; Claessens, Stijn; and Luc Laeven (2017). The Use and Effectiveness of Macroprudential Policies: New Evidence. Journal of Financial Stability, vol. 28, pp. 203–224.

Cohen, B. H. and Edwards, G. A. (2017). The new era of expected credit loss provisioning. BIS Quarterly Review.

Constâncio V. (2017). Resolving Europe's NPL Burden: Challenges and Benefits. Keynote speech at Bruegel Event: Tackling Europe's Non-performing Loans Crisis: Restructuring Debt, Reviving Growth, 3, February.

Council of the European Commission (2017). Report of the FSC Subgroup on non-performing loans. General Secretariat of the Council, 31 May.

de Lis, S. F., Pagés, J. M. and Saurina, J. (2001). Credit growth, problem loans and credit risk provisioning in Spain. In: Marrying the macro- and micro-prudential dimensions of financial stability, B. for International Settlements, ed., volume 1 of BIS Papers chapters, pages 331–353. Bank for International Settlements.

Diamond, D. W. and Dybvig, P. H. (1983). Bank runs, deposit insurance, and liquidity. The Journal of Political Economy, 91 (3): 401–419.

Draghi, M. (2017). European Banking Supervision Three Years on. Welcome Remarks by Mario Draghi, President of the ECB, at the Second ECB Forum on Banking Supervision, Frankfurt am Main, 7 November.

Drechsler, I., Savov, A. and Schnabl, P. (2018). Banking on Deposits: Maturity Transformation without Interest Rate Risk. NBER Working Paper No. w24582.

Eber, M. and Minoiu, C. (2016). How do banks adjust to stricter supervision? Technical report, mimeo.

European Banking Authority (2016). EBA report on the dynamics of non-performing exposures in the EU banking sector, 22 July.

European Banking Authority (2018). Risk dashboard data as of Q3 2018.

European Banking Authority (2019). EBA report on NPLs. Progress made and challenges ahead.

European Central Bank (2016). Stocktake of National Supervisory Practices and Legal Frameworks Related to NPLs, September.

European Central Bank (2017a). Guidance to banks on NPLs, March.

European Central Bank (2017b). Stocktake of National Supervisory Practices and Legal Frameworks Related to NPLs, July.

European Central Bank (2018). Addendum to the ECB Guidance to banks on nonperforming loans: supervisory expectations for prudential provisioning of non-performing exposures, March.

European Central Bank (2019). Communication on supervisory coverage expectations for NPEs, August.

European Systemic Risk Board (2017). Resolving non-performing loans in Europe, July.

European Systemic Risk Board (2019). Macro prudential approaches to non-performing loans, January.

Fell J., Grodzicki M., Martin R. and E. O'Brien (2016). Addressing market failures in the resolution of non-performing loans in the euro area. ECB Financial Stability Review – Special features, November

Gropp, R., Mosk, T., Ongena, S., and Wix, C. (2019). Banks Response to Higher Capital Requirements: Evidence from a Quasi-Natural Experiment. The Review of Financial Studies, 32(1), pp. 266-299.

Huizinga, H., Laeven, L. (2019). The Procyclicality of Banking: Evidence from the Euro Area. IMF Economic Review, Palgrave Macmillan; International Monetary Fund, 67(3):496-527.

Jiménez, G., Ongena, S., Peydro, J.-L. and Saurina, J. (2017). Macroprudential policy, countercyclical bank capital buffers, and credit supply: Evidence from the Spanish dynamic provisioning experiments. The Journal of Political Economy, 125(6):2126 - 2177.

Jiménez, G. and Saurina, J. (2006). Credit Cycles, Credit Risk, and Prudential Regulation. International Journal of Central Banking, 2(2).

Keeton, W. and C. Morris. (1987). Why Do Banks' Loan Losses Differ? Federal Reserve Bank of Kansas City, Economic Review, May, pp. 3-21.

Klein, N. (2013). Non-Performing Loans in CESEE: Determinants and Impact on Macroeconomic Performance. IMF Working Paper No. 13/72.

Kishan, R.P., and Opiela, T.P. (2000). Bank Size, Bank Capital, and the Bank Lending Channel. Journal of Money, Credit and Banking, 32: 121-141.

Laeven, L. and Majnoni, G. (2003). Loan loss provisioning and economic slowdowns: too much, too late? Journal of Financial Intermediation, 12(2):178 - 197.

Liu, C and Ryan, S.G. (2006). Income Smoothing over the Business Cycle: Changes in Banks' Coordinated Management of Provisions for Loan Losses and Loan Charge-Offs from the Pre-1990 Bust to the 1990s Boom. The Accounting Review, 81 (2): 421-441.

Medina Cas, S. and Peresa, I. (2016). What Makes a Good "Bad Bank" The Irish, Spanish and German Experience. European Economy - Discussion Papers 2015 - 036, Directorate General Economic and Financial Affairs (DG ECFIN), European Commission.

Nicoletti, A. (2018). The effects of bank regulators and external auditors on loan loss provisions. Journal of Accounting and Economics, 66 (81): 244-265.

Nkusu, M. (2011). Non-Performing Loans and Macrofinancial Vulnerabilities in Advanced Economies. IMF Working Paper 1/27

Norden, L. and Stoian, A. (2013). Bank earnings management through loan loss provisions: A double-edged sword? DNB Working Paper 404/December.

Ristolainen, K. (2018) Getting better? The effect of the single supervisory mechanism on banks' loan loss reporting and loan loss reserves. Bank of Finland Research Discussion Papers 11/2018.

Schivardi, F., Sette, E. and Tabellini, G. (2018). Credit Misallocation During the European Financial Crisis. EIEF Working Papers Series, Einaudi Institute for Economics and Finance (EIEF), No 1704.

Song, I. (2002). Collateral in loan classification and provisioning. IMF Working Paper 02/122.

Stamegna, C. (2019). Minimum loss coverage for non-performing loans. European Parliament Briefing.

Walter, J.R. (1991). Loan loss reserves. Economic Review, July/August.

Wheeler, P.B. (2019). Loan loss accounting and procyclical bank lending: The role of direct regulatory actions. Journal of Accounting and Economics, forthcoming.

Tables and Figures

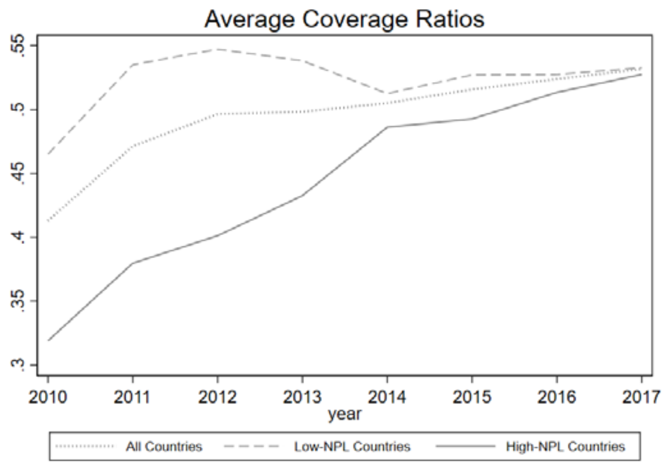


Figure 1: Average coverage ratio for all banks, banks from high-NPL countries, and banks from low NPL-countries. High-NPL countries (low-NPL countries) are defined as those with NPL/TA above (below) the sample mean. Data is winsorized at 2.5% and 97.5% (sample period: 2010–2017, source: authors' calculations).

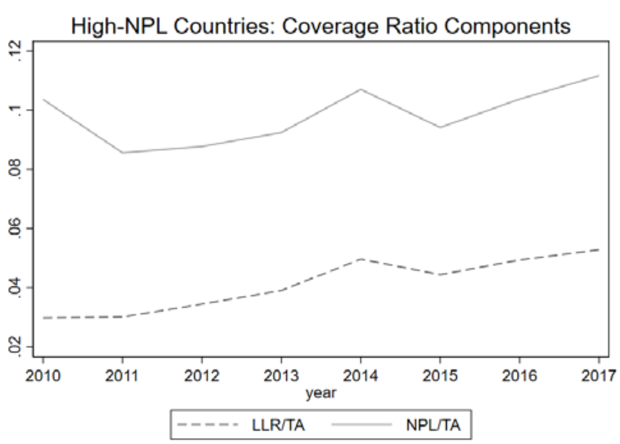


Figure 2: Average coverage ratio components (loan loss reserves and non-performing loans, scaled by total assets) for banks from high-NPL countries. High-NPL countries are defined as those with NPL/TA above the sample mean. Data is winsorized at 2.5% and 97.5% (sample period: 2010–2017, source: authors' calculations).

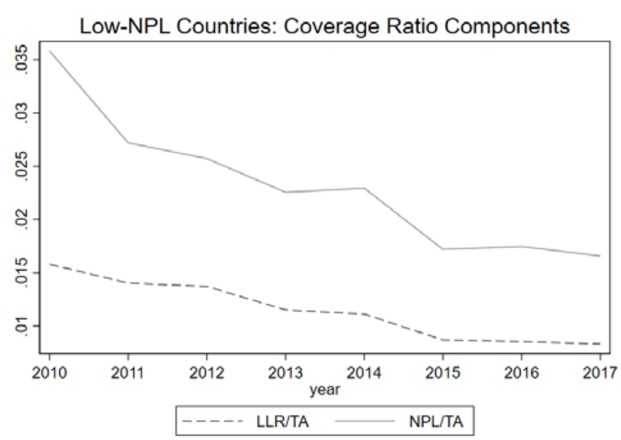


Figure 3: Average coverage ratio components (loan loss reserves and non-performing loans, scaled by total assets) for banks from low-NPL countries. Low-NPL countries are defined as those with NPL/TA below the sample mean. Data is winsorized at 2.5% and 97.5% (sample period: 2010–2017, source: authors' calculations).

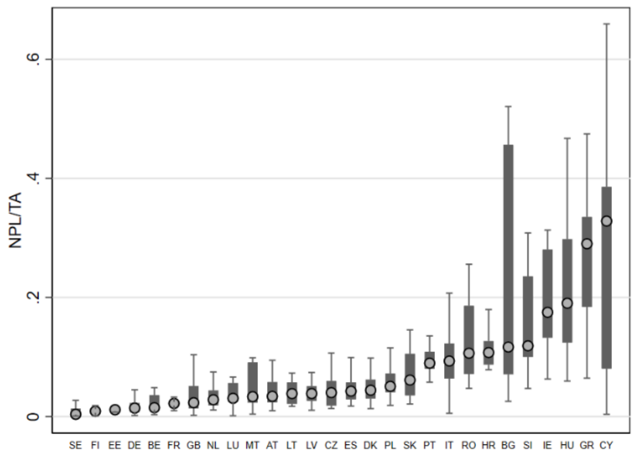


Figure 4: Boxplots of non-performing loans over total assets (NPL/TA) by country. Countries are ordered by median NPL/TA in ascending order. Data is winsorized at 2.5% and 97.5% by country (sample period: 2010–2017, source: authors' calculations).

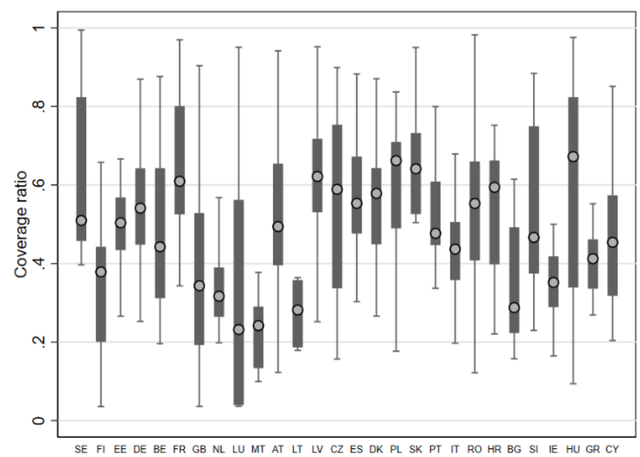


Figure 5: Boxplots of coverage ratios by country. Countries are ordered by median NPL/TA in ascending order. Data is winsorized at 2.5% and 97.5% by country. (sample period: 2010–2017, source: authors' calculations).

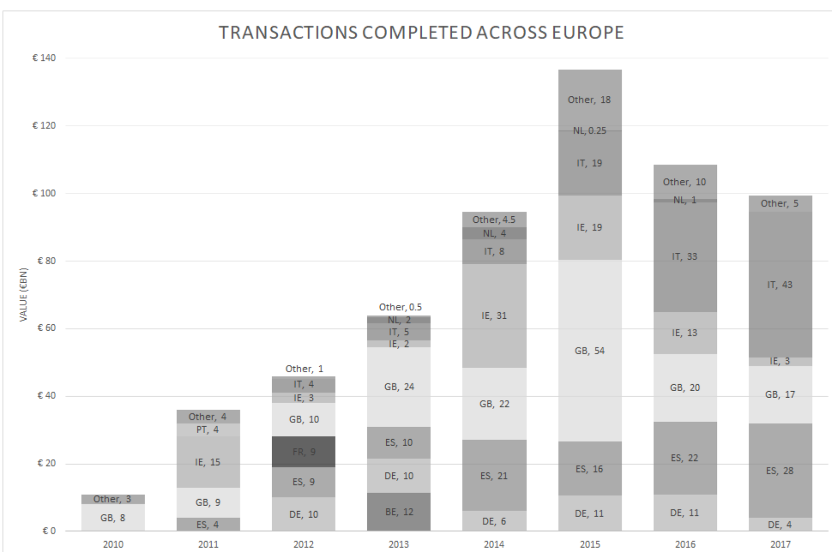


Figure 6: NPL secondary market transaction data (2010–2017, € billions, source: PwC)

Table 1: Correlations between coverage ratio and non-performing loans (NPL/TA), loan loss reserves (LLR/TA), and loan loss provisions (LLP/TA). Correlations with a * are significant at the 10% level.

	NPL/TA	NPL/TA $_{t-1}$	NPL/TA $_{t-2}$	LLR/TA	LLR/TA $_{t-1}$	LLR/TA $_{t-1}$	LLP/TA	LLP/TA $_{t-1}$	LLP/TA $_{t-2}$
Coverage ratio	-0.194*	-0.147*	-0.161*	0.010	0.057	0.080	-0.091	-0.081	-0.052
Coverage ratio$_{t-1}$	-0.195*	-0.165*	-0.165*	-0.010	0.056	0.093	-0.137*	-0.095	-0.059
Coverage ratio$_{t-2}$	-0.266*	-0.240*	-0.225*	-0.081	-0.023	0.045	-0.202*	-0.165*	-0.121

Table 2: Summary statistics for the baseline regression sample. Variables are winsorized at 2.5% and 97.5%.

	Mean	SD	Min	P10	P50	P90	Max	N
Bank Variables								
Coverage ratio	0.507	0.165	0.096	0.291	0.509	0.716	0.894	1845
LLR/TA	0.020	0.023	0.001	0.003	0.011	0.051	0.096	1845
NPL / TA	0.044	0.052	0.001	0.006	0.023	0.113	0.220	1845
Delta (NPL / TA)	-0.001	0.011	-0.028	-0.011	-0.002	0.010	0.041	1845
Gross loans / TA	0.649	0.133	0.294	0.465	0.665	0.809	0.886	1845
Gross loan growth	0.026	0.066	-0.127	-0.048	0.023	0.092	0.271	1845
log (Total Assets)	15.983	1.399	14.335	14.518	15.569	18.579	18.976	1845
Deposits / TA	0.660	0.162	0.280	0.392	0.709	0.827	0.907	1844
CET1	0.145	0.041	0.070	0.102	0.139	0.198	0.368	1843
ROAA	0.003	0.005	-0.016	0.000	0.002	0.007	0.018	1845
Cost-to-income ratio	0.654	0.118	0.364	0.498	0.662	0.794	0.944	1843
Institutional Variables								
Regulatory quality	1.430	0.437	0.148	0.711	1.687	1.817	2.047	1845
Rule of Law	1.398	0.547	-0.112	0.377	1.622	1.857	2.100	1845
Macroprudential Index	3.146	1.103	0.000	2.000	3.000	4.000	6.000	1845
Subcomponents of Macropru. Index								
Dynamic loan-loss provisioning	0.028	0.166	0.000	0.000	0.000	0.000	1.000	1845
Capital Surcharges on SIFI	0.394	0.489	0.000	0.000	0.000	1.000	1.000	1845
Levy/Tax on FI	0.778	0.416	0.000	0.000	1.000	1.000	1.000	1845
Loan-to-Value Ratio Caps	0.267	0.443	0.000	0.000	0.000	1.000	1.000	1845
NPL Secondary Market								
NPL secondary mkt / TA	0.003	0.006	0.000	0.000	0.002	0.006	0.117	1845
Business and Financial Cycle								
Real GDP growth rate	0.017	0.018	-0.091	0.003	0.019	0.029	0.252	1845
Unemployment rate	0.073	0.045	0.029	0.038	0.053	0.122	0.275	1845
House Price change (y-o-y)	0.023	0.041	-0.076	-0.045	0.028	0.073	0.076	1845
Stock Price change (y-o-y)	0.088	0.117	-0.252	-0.072	0.093	0.267	0.293	1845
Private sector credit / GDP	0.905	0.278	0.265	0.775	0.821	1.321	2.450	1845
Short-term interest rate	0.002	0.006	-0.007	-0.003	0.000	0.007	0.049	1845

Table 3: Preliminary analysis. The dependent variable is the coverage ratio. Only the constant and fixed effects at the bank and the country-year level are included.

	(1)	(2)	(3)
	Coverage ratio	Coverage ratio	Coverage ratio
Constant	0.507***	0.507***	0.507***
	(0.000)	(0.006)	(0.000)
Observations	1845	1845	1845
No. of banks	441	441	441
Adjusted R-squared	0.803	0.215	0.826
FE Bank	Yes	No	Yes
FE Country-year	No	Yes	Yes

Table 4: Micro-level regressions: without bank FE and baseline. The dependent variables are the coverage ratio, LLRs/TA, and NPLs/TA at the bank level. In columns 1-3 bank fixed effects are removed and replaced with bank-specific time invariant characteristics. In columns 4-5 bank fixed effects are introduced. Country-year dummies are included in each regression. Robust standard errors are clustered at the bank-level and reported in parentheses. Significance at the 1, 5 and 10% level is denoted by ***, **, and * respectively.

	Without Bank Fixed Effects			Baseline		
	(1)	(2)	(3)	(4)	(5)	(6)
	NPLs/TA	LLRs/TA	Coverage ratio	NPLs/TA	LLRs/TA	Coverage ratio
LLR/TA$_{t-1}$	1.231***		13.754***	1.056***		5.717***
	(0.077)		(0.973)	(0.126)		(0.581)
NPL / TA$_{t-1}$		0.337***	-5.806***		0.295***	-2.717***
		(0.023)	(0.494)		(0.035)	(0.347)
DELTA (NPL / TA)		0.309***	-1.100***		0.295***	-1.405***
		(0.042)	(0.415)		(0.039)	(0.267)
Gross loans / TA$_{t-1}$	0.021***	0.001	-0.099**	0.025*	0.010	-0.122
	(0.005)	(0.002)	(0.046)	(0.015)	(0.007)	(0.084)
Gross loan growth	-0.033***	0.001	0.091	-0.021*	0.008**	0.183***
	(0.010)	(0.004)	(0.062)	(0.011)	(0.004)	(0.040)
log (Total Assets)$_{t-1}$	-0.003***	0.000	-0.002	0.001	-0.001	0.052
	(0.001)	(0.000)	(0.008)	(0.008)	(0.002)	(0.033)
Deposits / TA$_{t-1}$	-0.009	0.002	-0.026	-0.026	0.003	0.061
	(0.006)	(0.003)	(0.049)	(0.016)	(0.005)	(0.072)
CET1$_{t-1}$	-0.035*	0.000	0.329*	0.019	0.011	0.306*
	(0.019)	(0.009)	(0.170)	(0.021)	(0.009)	(0.160)
ROAA$_{t-1}$	-0.985***	-0.021	-1.066	-0.635***	-0.273***	-0.885
	(0.201)	(0.097)	(1.206)	(0.205)	(0.088)	(0.666)
Cost-to-income ratio$_{t-1}$	-0.017**	0.000	-0.061	-0.021***	-0.002	-0.068**
	(0.007)	(0.003)	(0.050)	(0.007)	(0.003)	(0.034)
Significant	0.004	-0.001	-0.045*			
	(0.003)	(0.001)	(0.023)			
Listed	-0.002	0.000	0.044***			
	(0.002)	(0.001)	(0.016)			
Savings Mutual or Thrift	-0.003	0.000	0.015			
	(0.002)	(0.001)	(0.015)			
IFRS	0.006	-0.002	-0.034			
	(0.004)	(0.001)	(0.025)			
Well Reserved	0.004	0.018***	-0.086***			
	(0.005)	(0.002)	(0.022)			
Risky	0.046***	-0.009***	0.107***			
	(0.005)	(0.003)	(0.024)			
Loan-based	0.000	-0.001	-0.026*			
	(0.002)	(0.001)	(0.016)			
Large	0.002	0.001	0.019			
	(0.003)	(0.001)	(0.024)			
Deposit-based	0.004**	-0.001	-0.020			
	(0.002)	(0.001)	(0.023)			
Sound	0.004*	-0.002*	-0.050**			
	(0.002)	(0.001)	(0.021)			
Profitable	-0.004	0.004**	0.020			
	(0.004)	(0.002)	(0.027)			
Inefficient	-0.002	-0.002	-0.009			
	(0.003)	(0.001)	(0.020)			
Observations	1845	1845	1845	1845	1845	1845
No. of banks	441	441	441	441	441	441
Adjusted R-squared	0.922	0.93	0.561	0.956	0.968	0.853
Adjusted Within R-squared	0.778	0.802	0.441	0.319	0.520	0.157
FE Bank	No	No	No	Yes	Yes	Yes
FE Country-year	Yes	Yes	Yes	Yes	Yes	Yes

Table 5: Micro-level regressions: forward looking variable and high-NPL banks. The dependent variables are the coverage ratio and LLRs/TA at the bank level. In columns 1-2 we include the forward looking variable DELTA (NPL / TA)$_{t+1}$ as an independent variable. In columns 3-4 we include the dummy High NPL$_{t-1}$ to account for banks in the top decile of the NPL/TA ratio distribution, and its interactions with NPL/TA$_{t-1}$, ROAA$_{t-1}$, and CET$_{t-1}$. Country-year and bank dummies are included in each regression. Robust standard errors are clustered at the bank-level and reported in parentheses. Significance at the 1, 5 and 10% level is denoted by ***, **, and * respectively.

	Forward Looking		High NPL Banks	
	(1)	(2)	(3)	(4)
	LLRs/TA	Coverage ratio	LLRs/TA	Coverage ratio
LLR/TA$_{t-1}$		5.033***		5.620***
		(1.034)		(0.566)
NPL / TA$_{t-1}$	0.355***	-1.763***	0.331***	-3.403***
	(0.023)	(0.487)	(0.034)	(0.397)
DELTA (NPL / TA)	0.330***	-1.221***	0.310***	-1.451***
	(0.036)	(0.,313)	(0.037)	(0.263)
Gross loans / TA$_{t-1}$	-0.003	-0.131*	0.011	-0.118
	(0.005)	(0.068)	(0.007)	(0.085)
Gross loan growth	0.006	0.085*	0.007*	0.188***
	(0.005)	(0.052)	(0.004)	(0.039)
log (Total Assets)$_{t-1}$	0.000	-0.021	-0.001	0.051
	(0.002)	(0.030)	(0.002)	(0.034)
Deposits / TA$_{t-1}$	0.009*	0.012	0.004	0.036
	(0.005)	(0.080)	(0.005)	(0.072)
CET1$_{t-1}$	0.015	0.315*	0.003	0.329**
	(0.011)	(0.164)	(0.009)	(0.159)
ROAA$_{t-1}$	-0.101	0.479	-0.279***	0.068
	(0.085)	(0.812)	(0.089)	(0.837)
Cost-to-income ratio$_{t-1}$	-0.005*	-0.072*	-0.001	-0.065**
	(0.003)	(0.040)	(0.003)	(0.033)
DELTA (NPL / TA)$_{t+1}$	0.016	1.237***		
	(0.028)	(0.255)		
High NPL Dummy$_{t-1}$			-0.005	-0.109**
			(0.006)	(0.050)
*High NPL dummy$_{t-1}$ * NPL/TA$_{t-1}$*			-0.080*	0.996***
			(0.044)	(0.372)
*High NPL dummy$_{t-1}$ * CET1$_{t-1}$*			0.132***	0.227
			(0.042)	(0.292)
*High NPL dummy$_{t-1}$ * ROAA$_{t-1}$*			-0.152	-2.150**
			(0.176)	(1.073)
Observations	1251	1251	1845	1845
No. of banks	348	348	441	441
Adjusted R-squared	0.977	0.878	0.969	0.856
Adjusted Within R-squared	0.615	0.112	0.543	0.171
FE Bank	Yes	Yes	Yes	Yes
FE Country-year	Yes	Yes	Yes	Yes

Table 6: Micro-macro regressions: baseline. The dependent variables are the coverage ratio and LLRs/TA at the bank level. High-NPL countries are defined as countries with NPL/TA above the sample mean. Bank, business cycle and financial cycle controls as well as bank and time dummies are included in each regression. Robust standard errors are clustered at the bank-level and reported in parentheses. Significance at the 1, 5 and 10% level is denoted by ***, **, and * respectively.

	(1) NPLs/TA	(2) LLRs/TA	(3) Coverage ratio	(4) NPLs/TA	(5) LLRs/TA	(6) Coverage ratio
Regulatory Quality	0.008	-0.004*	-0.014	0.004	-0.003	-0.017
	(0.006)	(0.003)	(0.036)	(0.007)	(0.003)	(0.038)
Rule of Law	-0.015	0.004	0.014	-0.014	0.006	0.029
	(0.009)	(0.004)	(0.034)	(0.009)	(0.004)	(0.033)
Macroprudential Index	0.001	0.001*	0.012**			
	(0.001)	(0.000)	(0.005)			
Dynamic loan-loss provisioning				-0.019***	0.006	0.069**
				(0.006)	(0.004)	(0.027)
Capital Surcharges on SIFI				0.000	0.001	0.004
				(0.003)	(0.001)	(0.012)
Levy/Tax on Financial Institutions				0.003	0.002	0.020*
				(0.002)	(0.001)	(0.011)
Loan-to-Value Ratio Caps				0.002	0.001	-0.006
				(0.003)	(0.002)	(0.013)
High NPL Country Dummy	0.033***	0.005*	-0.006	0.022***	0.004*	-0.034*
	(0.005)	(0.002)	(0.023)	(0.005)	(0.003)	(0.020)
*High NPL Country * Regulatory Quality*	0.000	-0.003	0.021	0.010	-0.002	0.061*
	(0.006)	(0.002)	(0.031)	(0.007)	(0.003)	(0.033)
*High NPL Country * Rule of Law*	-0.012**	-0.001	-0.040*	-0.018***	-0.001	-0.065***
	(0.005)	(0.002)	(0.024)	(0.007)	(0.003)	(0.025)
*High NPL Country * Macroprudential Index*	-0.002	0.000	-0.005			
	(0.001)	(0.000)	(0.005)			
*High NPL Country * Dynamic LLP*				0.007	0.002	-0.057*
				(0.007)	(0.003)	(0.033)
*High NPL Country * Cap. Sur (SIFI)*				-0.007***	0.001	-0.026***
				(0.002)	(0.001)	(0.009)
*High NPL Country * Levy on FI*				0.000	0.001	-0.007
				(0.004)	(0.001)	(0.012)
*High NPL Country * LTV Caps*				0.009**	0.001	0.020
				(0.004)	(0.002)	(0.013)
Observations	1845	1845	1845	1845	1845	1845
No. of banks	441	441	441	441	441	441
Adjusted R-squared	0.954	0.962	0.86	0.956	0.963	0.861
Adjusted Within R-squared	0.550	0.668	0.228	0.569	0.675	0.233
Bank Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
Business and Financial Cycle Control Variables	Yes	Yes	Yes	Yes	Yes	Yes
FE Bank	Yes	Yes	Yes	Yes	Yes	Yes
FE Year	Yes	Yes	Yes	Yes	Yes	Yes

Table 7: NPL Secondary Market Transactions. NPL secondary market transaction/TA measures the share of NPL transactions over the total banking assets at the country level. The dependent variables are the coverage ratio, LLRs/TA, and NPLs/TA at the bank level. Bank, business cycle and financial cycle controls, as well as bank and time dummies are included in each regression. Robust standard errors are clustered at the bank-level and reported in parentheses. Significance at the 1, 5 and 10% level is denoted by ***, **, and * respectively.

	(1) NPLs/TA	(2) LLR/TA	(3) Coverage ratio	(4) NPLs/TA	(5) LLR/TA	(6) Coverage ratio	(7) NPLs/TA	(8) LLR/TA	(9) Coverage ratio
NPL Secondary Market Transactions / TA	-0.007	0.160**	-0.295						
	(0.144)	(0.069)	(0.472)						
Medium NPL Secondary Mkt				0.000	0.001**	0.010	0.001	0.002**	0.006
				(0.002)	(0.001)	(0.008)	(0.002)	(0.001)	(0.010)
Large NPL Secondary Mkt				0.005**	0.002***	0.021***	0.004**	0.001*	0.012
				(0.002)	(0.001)	(0.008)	(0.002)	(0.001)	(0.009)
High NPL Country							0.012***	0.001	-0.053***
							(0.002)	(0.001)	(0.009)
*High NPL Country * Medium NPL Secondary Mkt*							0.001	0.000	0.007
							(0.002)	(0.001)	(0.010)
*High NPL Country * Large NPL Secondary Mkt*							0.003	0.002**	0.018**
							(0.002)	(0.001)	(0.009)
Observations	1845	1845	1845	1845	1845	1845	1845	1845	1845
No. of banks	441	441	441	441	441	441	441	441	441
Adjusted R-squared	0.947	0.961	0.854	0.948	0.961	0.854	0.953	0.962	0.860
Adjusted Within R-squared	0.482	0.663	0.193	0.487	0.660	0.196	0.535	0.665	0.227
Bank Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Business and Financial Cycle Control Variables	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE Bank	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
FE Year	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

APPENDIX A1- Additional tables

Table A.1: Sample composition and average coverage ratio, LLR/TA and NPL/TA ratios by country.

Country Code	No. Observations	No. Banks	Avg. Coverage Ratio	Avg. LLR/TA	Avg. NPL/TA
AT	69	19	53%	2%	4%
BE	20	4	47%	1%	2%
BG	11	3	32%	5%	13%
CY	8	2	41%	9%	20%
CZ	20	4	63%	3%	4%
DE	938	232	54%	1%	2%
DK	48	9	55%	4%	7%
ES	74	17	56%	2%	5%
FI	15	6	34%	0%	1%
FR	67	18	63%	1%	2%
GB	128	30	36%	1%	4%
GR	24	5	42%	9%	21%
HR	8	2	57%	6%	11%
HU	9	3	66%	9%	16%
IE	13	3	36%	6%	16%
IT	252	48	44%	5%	10%
LT	4	2	27%	1%	4%
LU	10	2	32%	1%	4%
LV	8	3	54%	2%	4%
MT	9	3	27%	1%	4%
NL	31	6	31%	1%	4%
PL	32	6	59%	3%	6%
PT	20	5	54%	5%	10%
RO	8	2	56%	4%	8%
SE	4	2	72%	1%	1%
SI	9	3	66%	6%	9%
SK	6	2	65%	3%	4%
Total	1845	441			

Table A.2: Correlation matrix of the independent and dependent variables in our baseline (micro and micro-macro) analyses. Correlations with a * are significant at the 10% level.

	Cov. ratio	*LLR / TA*	*NPL / TA*	*DELTA (NPL / TA)*	*Gross loans / TA*	*Gross loan growth*	*log (TA)*	*Dep / TA*	*CET1*	*ROAA*	*Cost-to-inc. ratio*	*Reg. Quality*	*Rule of Law*	*Macro-pru, Index*	*Dyn. LLP*	*Cap.Sur (SIFI)*	*Levy / Tax on FI*	*LTV Ratio Caps*	*NPL sec. mkt*
Cov. ratio	1																		
LLR/TA	0.019	1																	
NPL / TA	-0.227*	0.938*	1																
DELTA (NPL / TA)	-0.159*	0.259*	0.342*	1															
Gross loans/ TA	-0.116*	0.209*	0.222*	0.067*	1														
Gross loan growth	0.065*	-0.263*	-0.301*	-0.087*	0.011	1													
log (Total Assets)	-0.116*	0.092*	0.107*	-0.006	-0.073*	-0.136*	1												
Dep. / TA	0.118*	-0.192*	-0.236*	-0.163*	0.087*	0.218*	-0.498*	1											
CET1	0.096*	-0.145*	-0.167*	-0.126*	-0.158*	0.050*	-0.177*	0.186*	1										
ROAA	0.128*	-0.198*	-0.270*	-0.358*	-0.038	0.301*	-0.044*	0.073*	0.229*	1									
Cost-to-income ratio	-0.027	-0.103*	-0.096*	-0.008	0.029	-0.106*	-0.082*	0.135*	-0.181*	-0.411*	1								
Regulatory Quality	0.053*	-0.655*	-0.630*	-0.201*	-0.049*	0.088*	-0.156*	0.294*	0.160*	-0.005	0.159*	1							
Rule of Law	0.093*	-0.630*	-0.623*	-0.236*	-0.080*	0.055*	-0.047*	0.184*	0.130*	0.022	0.125*	0.923*	1						
Macropru. Index	0.232*	0.018	-0.059*	-0.170*	-0.014	0.085*	-0.156*	0.133*	0.191*	0.110*	0.031	-0.084*	-0.166*	1					
Dynamic LLP	0.023	0.087*	0.064*	-0.033	0.008	0.015	0.106*	-0.013	-0.060*	0.100*	-0.127*	-0.242*	-0.186*	0.098*	1				
Cap. Sur (SIFI)	0.121*	-0.141*	-0.174*	-0.179*	0.015	0.114*	-0.112*	0.181*	0.233*	0.092*	0.070*	0.170*	0.006	0.684*	-0.136*	1			
Levy/Tax on FI	0.264*	-0.229*	-0.297*	-0.224*	-0.105*	0.003	-0.224*	0.204*	0.119*	-0.073*	0.223*	0.269*	0.281*	0.498*	-0.205*	0.246*	1		
LTV Ratio Caps	-0.006	0.434*	0.398*	-0.014	0.001	-0.033	0.115*	-0.196*	-0.029	0.178*	-0.183*	-0.478*	-0.415*	0.394*	0.212*	0.108*	-0.082*	1	
NPL sec. mkt	-0.085*	0.275*	0.295*	-0.052*	0.070*	-0.043*	0.074*	-0.067*	0.004	0.022	-0.019	-0.165*	-0.223*	0.085*	0.023	0.048*	-0.005	0.260*	1

Table A.3: Micro-level regressions: robustness. The dependent variables are the coverage ratio, LLRs/TA, and NPLs/TA at the bank level. Within the explanatory variables, NPL/TA is replaced with NPLs over gross loans (NPL/GL), the CET1 ratio is replaced by the Tier 1 Capital ratio, and ROAA is replaced with the return-on-equity (ROAE). Robust standard errors are clustered at the bank-level and reported in parentheses. Significance at the 1, 5 and 10% level is denoted by ***, **, and * respectively.

	(1) NPLs/TA	(2) LLRs/TA	(3) Coverage ratio
LLR/TA$_{t-1}$	1.035***		4.807***
	(0.125)		(0.515)
NPL/Gross Loans$_{t-1}$		0.211***	-1.739***
		(0.027)	(0.199)
DELTA (NPL /GL)		0.173***	-1.078***
		(0.025)	(0.156)
Gross Loans/TA$_{t-1}$	0.025*	0.026***	-0.266***
	(0.015)	(0.006)	(0.090)
Gross loan growth	-0.020*	0.017***	0.136***
	(0.011)	(0.005)	(0.043)
log (Total Assets)$_{t-1}$	0.001	0.003	0.033
	(0.008)	(0.002)	(0.035)
Deposits/TA$_{t-1}$	-0.023	0.005	0.058
	(0.015)	(0.005)	(0.073)
Tier 1 Capital$_{t-1}$	0.023	0.014	0.227
	(0.023)	(0.010)	(0.167)
ROAE$_{t-1}$	-0.055***	-0.021***	-0.115**
	(0.016)	(0.006)	(0.051)
Cost-to-income ratio$_{t-1}$	-0.021***	0.001	-0.085**
	(0.007)	(0.003)	(0.033)
Observations	1842	1842	1842
No. of banks	441	441	441
Adjusted R-squared	0.956	0.966	0.853
Adjusted Within R-squared	0.321	0.491	0.155
FE Bank	Yes	Yes	Yes
FE Country-year	Yes	Yes	Yes

Table A.4: Shapley decomposition

Panel A

Variable	Value	In percentage
LLR/TA_{t-1}	0.199	38.71%
NPL / TA_{t-1}	0.187	36.48%
DELTA (NPL / TA)	0.006	1.23%
Gross loans / TA $_{t-1}$	0.012	2.40%
Gross loan growth	0.004	0.69%
log (Total Assets)$_{t-1}$	0.009	1.74%
Deposits / TA$_{t-1}$	0.004	0.77%
CET1$_{t-1}$	0.001	0.28%
ROAA$_{t-1}$	0.003	0.55%
Cost-to-income ratio$_{t-1}$	0.001	0.16%
Group: Macro	0.087	16.99%
TOTAL	0.513	100.00%

Panel B

Variable	Value	In percentage
Institutional Variables		
Regulatory Quality	0.003	0.65%
Rule of Law	0.006	1.18%
Macroprudential Index	0.027	5.25%
Business and Financial Cycle		
GDP growth$_{t-1}$	0.003	0.53%
Unemployment$_{t-1}$	0.002	0.41%
House Price change y-o-y$_{t-1}$	0.009	1.73%
Stock Price change y-o-y$_{t-1}$	0.001	0.23%
Private credit to GDP$_{t-1}$	0.025	4.91%
Short term interest rate$_{t-1}$	0.005	1.06%
Group: Micro	0.431	84.05%
TOTAL	0.513	100.00%