Information Gaps and Price Efficiency: The Role of Social Media

Jonathan Brogaard
University of Utah
Email: jonathan.brogaard@eccles.utah.edu

Jan Hanousek, Jr.
University of Memphis
Email: jan.hanousek@memphis.edu

Dominik Svoboda
University of Memphis
Email: dsvoboda@memphis.edu

Velma Zahirovic-Herbert
University of Memphis
Email: Velma.Zahirovic-Herbert@memphis.edu

Abstract

We study the role of social media in financial markets in a setting with well-defined information frictions. Using exogenous shocks from natural disasters and a delay in official reporting, we identify periods in which investors lack reliable information about local fundamentals. We show that in the absence of coverage, municipal bond yields exhibit an initial reaction followed by a predictable correction as information becomes available. Exploiting the staggered introduction of location-based social media, we find that these corrections disappear when decentralized information is present. Our results demonstrate that social media improves price efficiency by expanding information availability when traditional sources are limited.

Keywords: social media, information friction, price efficiency

JEL Codes: G14, D83, G12, L82

The research was supported by GAČR grant No. 26-23313S. The usual disclaimer applies.

1. Introduction

There is a growing literature studying the role of social media in financial markets. A central theme in this work is that, while social media can contain economically relevant information, it often generates distortions rather than positive externalities. Social media facilitates coordination among investors and can amplify destabilizing price movements (Bradley et al., 2024), fosters echo chambers (Cookson et al., 2023) and disagreement (Cookson et al., 2020), and is increasingly subject to manipulation through bots (Hanousek et al., 2026; Bradley et al., 2026). Taken together, this evidence suggests that the optimal choice for a rational investor is to avoid social media altogether and instead rely on traditional media and official information sources.

This conclusion, however, overlooks a more fundamental friction. Traditional media provides incomplete coverage of economically relevant events. Attention is concentrated on a narrow set of entities and shocks, leaving large segments of the market effectively uncovered (Engelberg & Parsons, 2011). When such coverage gaps arise, investors face severe information constraints, which can lead to delayed price discovery and mispricing (Hung et al., 2015). In these environments, the relevant comparison is not between high-quality traditional news and noisy social media, but between some information and no information at all.

We study this question in a setting with a well-defined gap in official information and clear variation in local information availability. We exploit federally declared natural disasters as plausibly exogenous shocks to the fundamentals of municipal bond issuers. Following the shock, investors face a well-defined information gap: while the Federal Emergency Management Agency (FEMA) collects data on damages, there is no official information about the magnitude of the shock for approximately 30 days. During this informational black-out period, information is determined by media coverage, which is sparse and highly selective, leaving many affected counties without coverage. We aggregate bond yields at the monthly frequency, which maps directly into this structure and partitions the data into distinct information regimes following the realization of the shock: a period of limited and uneven information, and the subsequent arrival of official estimates. This setting provides a transparent framework to study how variation in information availability affects price adjustment.

We observe that the majority of locations affected by federally declared emergencies were not covered at all by traditional news publications. We observe a clear pattern consistent with mispricing under limited information. When the disaster happens, yields rise significantly, reflecting heightened perceived risk. However, after the official estimates from FEMA are released, we observe a correction, as yields partially reverse. This pattern is consistent with an initial overreaction under limited information and a later correction as more reliable information becomes available

We next examine whether social media helps mitigate these information gaps and frictions by introducing systematic variation in information availability within this setting. We focus on Reddit, which is organized into topic-specific communities ("subreddits"). Our primary focus is on location-specific communities, which cover specific counties or cities and can provide real-time information about local conditions. Crucially, the presence of these communities varies across counties and emerges gradually over time, generating a staggered rollout of social media coverage. We exploit this staggered introduction in a difference-in-differences framework to identify the causal effect of social media on price discovery. In counties with local Reddit coverage, prices appear to adjust more fully during the 30-day information gap, reducing the later response to official estimates. As a result, the subsequent arrival of FEMA data no longer leads to a significant price correction. These results show that social media attenuates mispricing by supplying decentralized local information precisely when traditional and official sources are absent.

Our findings highlight that the role of social media depends critically on the surrounding information environment. While prior work emphasizes the distortions introduced by social media, our results show that its informational content becomes valuable when traditional and official sources are absent. In our setting, the key margin is not the speed of price adjustment, but whether prices reflect the available information. When investors lack information, prices following the shock reflect uncertainty and are subsequently corrected once reliable data is released. In contrast, when social media provides even imperfect signals about local conditions, prices appear to adjust earlier, and the later response to official estimates is substantially smaller. Social media thus appears to improve market outcomes by reducing information gaps rather than merely accelerating price discovery.

Our results do not imply that social media substitutes for traditional or official information sources, nor that it is uniformly reliable. Instead, its value arises precisely in environments where these sources provide limited or no coverage.[1] In such settings, social media acts as a decentralized information channel that complements traditional media by expanding the set of observable events. More broadly, our empirical design mirrors a simple theoretical environment with an exogenous shock, a period of limited information, and a subsequent public signal. This structure allows us to isolate how alternative information sources affect equilibrium pricing. We show that market

[1] Even in the stock market, there are numerous examples of limited or no coverage. Numerous firms are without any single sell-side analysts or lose coverage after previously having one (Mola et al., 2013). Thus, even in today's age, majority of the attention is concentrated on the small subset of large stocks, leaving many with information gaps analogous to the ones we are analyzing in municipal bond setting.

efficiency depends not only on the precision of information, but also on its availability. When the alternative is no information, even noisy signals can improve price accuracy and efficiency.

Our paper contributes to several strains of literature. Firstly, this paper contributes to the growing literature on social media and financial markets. Prior work shows that social media contains economically relevant information but often emphasizes its distortive effects, including coordinated trading, sentiment-driven price pressure, and the formation of echo chambers (e.g., Tetlock, 2007; Gurun & Butler, 2012; Barber et al., 2024; Cookson et al., 2024; Goldstein et al., 2025). This body of work suggests that social media may amplify noise or disagreement, raising the question of why sophisticated investors would rely on it. We provide a complementary perspective by showing that the value of social media depends critically on the surrounding information environment. In settings where traditional media and official sources provide limited coverage, social media can act as a decentralized information intermediary that expands the set of available signals. Using exogenous shocks and staggered variation in social media coverage, we show that these signals are incorporated into prices and eliminate predictable corrections that arise in their absence. Our results thus reconcile the informational content of social media with its documented distortions by highlighting its role in environments characterized by incomplete information rather than excess noise.

Our paper also contributes to the literature on information frictions and price discovery. A large body of theoretical work shows that when information about fundamentals is incomplete or dispersed, prices may deviate from fundamentals and adjust as new signals arrive (e.g., Grossman and Stiglitz, 1980; Diamond and Verrecchia, 1981; Morris and Shin, 2002; Vives, 2008; Goldstein et al., 2013). Empirically, however, isolating the role of information availability is challenging due to endogenous information flows and confounding variation in fundamentals. We address this challenge by exploiting a setting that closely mirrors a canonical information environment, combining exogenous shocks, a well-defined information gap, and a delayed public signal. Within this framework, we exploit staggered variation in the availability of decentralized information through the rollout of social media coverage. Our findings show that expanding the information set—even with imperfect signals—can substantially improve price accuracy by eliminating predictable revisions. More broadly, we provide direct empirical evidence that market efficiency depends not only on the precision of information, but also on its availability.

The study is organized as follows. We outline the theoretical framework in Section 2. We describe the data and the empirical methodology set up in Section 3. We show the results of our analysis in Section 4. Lastly, Section 5 concludes.

2. Framework and predictions

Market efficiency depends fundamentally on the availability and incorporation of information. In standard asset pricing frameworks, prices reflect investors' expectations about fundamentals, which in turn depend on the information available to market participants (Fama, 1970). When information is complete and publicly available, prices adjust to reflect underlying fundamentals. However, a large theoretical literature emphasizes that when information is incomplete, dispersed, or costly to acquire, prices may deviate from fundamentals and only gradually incorporate new information (e.g., Grossman and Stiglitz, 1980; Vives, 2008). In such environments, the efficiency of prices depends not only on the precision of available signals, but also on whether relevant information is observed at all. When economically important events are not covered by traditional information intermediaries, investors face fundamental uncertainty about the magnitude and persistence of shocks, which can lead to mispricing.

Municipal bonds provide a useful setting for our analysis because repayment capacity is closely tied to issuing municipality. Natural disasters serve as plausible exogenous events that can affect local tax base, spending needs, and expected fiscal pressure (Addoum & Ortiz-Bobea, 2020; Painter, 2020). The severity of those effects is not immediately observable to outside investors. Instead, investors must infer local conditions from the information environment, which might be sparse, prior to any official estimates that provide more complete information.

Our empirical framework resembles a rational expectations setting with delayed public information and heterogeneous private signals. Consider a discrete-time framework in which an exogenous shock to fundamentals is realized in the initial period, but its magnitude is not immediately observed by investors. Instead, agents receive imperfect and potentially heterogeneous signals about the shock, and a precise public signal is released with delay. In such settings, prices reflect the aggregation of available signals prior to the arrival of the public signal, and may be revised after more precise information becomes available (e.g., Diamond and Verrecchia, 1981; Morris and Shin, 2002; Vives, 2008). Our setting provides an empirical analogue to this environment. Federally declared natural disasters represent plausibly exogenous shocks, while the delay in FEMA reporting generates a period in which investors observe the occurrence of the shock but lack precise information about its magnitude. During this period, information is determined by available signals, which vary across locations depending on media coverage. The subsequent release of FEMA estimates corresponds to the arrival of a public signal that resolves uncertainty. While we do not observe investor attention to subreddit content directly, we use the existence of location-based online communities as a proxy for the local availability of decentralized information. We use the rational-expectations framework primarily to motivate the timing of information arrival, while interpreting the later predictable reversal more cautiously as delayed adjustment under limited information or the resolution of uncertainty once more reliable public signals arrive.

Within this framework, social media naturally enters as a source of decentralized and heterogeneous signals about the underlying shock. In the absence of comprehensive traditional media coverage, investors may rely on alternative information channels that provide noisy but informative signals about local conditions. We focus on Reddit, which is organized into decentralized system of communities. We focus on location-specific, i.e., county and city, communities that generate real-time, ground-level information. The presence of these communities varies across counties and evolves over time, creating a staggered introduction of additional signals into the information environment. This variation allows us to move beyond a setting with limited information to one in which investors observe an expanded set of signals prior to the arrival of the public signal. In the language of standard information aggregation models, social media increases the information set available to market participants, potentially improving the extent to which prices reflect fundamentals before uncertainty is resolved. Our empirical design exploits this staggered expansion of the information set to identify how the availability of additional, decentralized signals affects equilibrium pricing. We explicitly outline the framework and the predictions in Figure 1.

Insert Figure 1

3. Data and Empirical Design

3.1 Institutional setting and data construction

We combine transaction-level municipal bond data with exogenous shocks to local economic conditions and measures of information coverage. Municipal bond transactions are obtained from the Municipal Securities Rulemaking Board (MSRB), which provides detailed information on secondary market trades, including yields, traded volume, coupon, maturity, and trade type. To ensure data quality, we follow standard filters in the literature (Harris & Piwowar, 2006; Green et al., 2007; Gao et al., 2020), excluding bonds with maturity longer than 100 years, variable coupon rates, or bonds with fewer than ten transactions over the sample period, as well as trades with implausible yield values. Bond characteristics and issuer location are obtained from Bloomberg, which allows us to map each bond to its issuing county.

We identify exogenous shocks using data from FEMA's Disaster Declarations Summaries, which provide the timing and geographic scope of federally declared natural disasters.[2] To align the timing of the shock with our event structure, we restrict the sample to disasters declared within seven days of the incident start, allowing for limited administrative delays.[3] We further focus on the most common large-scale disaster types—hurricanes, severe storms, floods, wildfires, and tornadoes—to ensure sufficient coverage and comparability across events. The resulting sample contains 90 disasters across these categories.

To capture variation in information availability from traditional media, we use data from RavenPack, which provides location-specific news coverage. Following prior literature (e.g., You et al., 2018; Bushman & Pinto, 2023; Cookson et al., 2024), we identify whether a county receives

[2] The FEMA data for damage estimates, which provide full information for the investors are available through the FEMA Individual and Household Program (IHP), reported with 30-day delay. The IHP contains applicant-level records for disaster relief support and includes verified locations at the county-state level.

[3] Let us note that there are some events, where disaster is declared more than two months following the reported end of the event. This delay in disaster declarations are typically related to political considerations or to help with the recovery of the area. In these events, the 30 day information impact by FEMA is unlikely to significantly impact the market, given the long delay and other informational spillovers.

coverage in the 30 days following a disaster declaration.[4,5] A county is classified as covered if at least one article explicitly references the disaster and the affected county or a city within it. This measure allows us to identify systematic gaps in traditional media coverage across affected locations.

The final dataset contains approximately 92,000 bonds mapped to affected and unaffected counties across all events. We report the distribution of covered and non-covered events across the timeline of our sample in Figure 2. Importantly, a large share of affected counties receive no coverage, creating a setting in which investors must form beliefs in the absence of traditional information intermediaries.

Insert Figure 2

We further observe substantial cross-sectional variation in media coverage: while large and prominent counties receive consistent attention, many affected regions are not covered at all. This variation in coverage plays a central role in our empirical design. We report the frequency of coverage of all locations following disasters in Figure 3.

Insert Figure 3

3.2 Event time aggregation

---

[4] We only focus on the newspaper coverage during this period after emergency was declared. Articles written prior to the start of the event are not considered, as they cannot provide information about the impact of the disaster. Similarly, newspaper coverage after FEMA starts reporting estimates of damages are also not considered, as the information will be less salient than the FEMA coverage.

[5] Given the potential discrepancy between start of the event and declaration of emergency, we further consider any articles written during this period as coverage of the event. For example, if FEMA considers start of the event on Monday, but emergency was declared on Friday, we consider the location covered if there are articles mentioning the disaster and location written from Monday to Friday on top of the 30 days following Friday. Our results are robust to the exclusion of these articles.

We construct the empirical counterpart to the information structure described in Section 2 by aggregating bond yields into 30-day event-time intervals. This choice is motivated by the institutional setting: the delay in FEMA reporting generates a natural 30-day window in which no official information about the magnitude of the shock is available. For each bond, we compute the average secondary market yield across all customer buy transactions within each 30-day window, weighted by par value traded, following standard practice in the municipal bond literature (e.g., Gao et al., 2020). We define event time $t = 0$ as the 30 days following the disaster declaration, with $t = -1$ corresponding to the preceding 30 days, and subsequent periods defined analogously.

Our primary outcome variable is the change in yields across adjacent event-time periods, $\Delta Yield_t = Yield_t - Yield_{t-1}$. This transformation plays a central role in our empirical design. By focusing on within-bond changes over short horizons aligned with the event structure, we difference out time-invariant bond and location characteristics, as well as highly persistent local conditions. As a result, identification relies on variation in yield changes around the event, rather than cross-sectional differences in levels. Importantly, this substantially reduces the need for extensive control variables. Many commonly used controls in the municipal bond literature, such as county-level economic conditions (e.g., Gao et al., 2020), are measured at annual or quarterly frequency and exhibit little variation at the monthly horizon. Including such variables in a differenced specification can introduce measurement error without improving identification. In contrast, our specification, combined with county and event-time fixed effects, isolates the effect of information availability on price adjustment in a transparent and parsimonious way.

3.3 Empirical design

We estimate the effect of information availability on price dynamics using a stacked difference-in-differences design. For each disaster, we construct an event-level panel that includes all bonds issued by counties within affected states and stack these panels across events. This approach ensures that identification is driven by within-event variation in exposure and information availability, while allowing for flexible comparisons across disasters. Our baseline specification relates changes in bond yields to event indicators around the disaster:

$$\Delta Yield_{i,c,g,t} = \beta_0 + \beta_1 CoveragepGap_{c,g,t} + \beta_2 Affected_{c,g,t} + \beta_3 CoveragepGap_{c,g,t} \times Affected_{c,g,t} + Bond\ Controls + Fixed\ Effects + \epsilon_{i,c,g,t}, \qquad (1)$$

where the dependent variable $\Delta Yield_{i,c,g,t}$ is defined as the change in yield for bond *i* in county *c* related to disaster *g* during a 30-day window *t*.[6] The variable *Affected* equals one if the county was affected in the disaster, namely event *g* occurs in county *c* during time *t=0*. Similarly, Coverage Gap captures the absence of location-specific media coverage, and is equal to one if the county is not explicitly mentioned in public news articles mentioning the event during the coverage window t=0, and 0 otherwise. Thus, the interaction term $Affected \times CoverageGap$ identifies counties that experience a disaster but do not receive location-specific media coverage. This allows us to compare counties exposed to the same underlying shock but facing different levels of information availability. We further specify simplified model, where we restrict the sample to only affected counties.

$$\Delta Yield_{i,c,g,t} = \beta_0 + \beta_1 CoveragepGap_{c,g,t} + Controls + \delta_{g,t} + \gamma_c + \epsilon_{i,c,g,t}, \qquad (2)$$

[6] Where $\Delta Yield_{i,c,g,t} = Yield_{i,c,g,t} - Yield_{i,c,g,t-1}$.

Both specifications include county fixed effects, which absorb time-invariant local characteristics, and event-time fixed effects, which capture common shocks across disasters. We additionally include credit rating fixed effects to control for differences in baseline credit risk across securities. Given the short event-time horizon and the use of yield differences, this specification already absorbs most sources of cross-sectional and temporal variation. We further restrict the sample to bonds that are traded in all periods, to create a consistent cross section and to mitigate the the risk that our results are artifacts of changing sample composition or the entry/exit of distressed issues.[7] As discussed in Section 3.2, many commonly used control variables in the municipal bond literature vary only at low frequencies and contribute little additional identifying variation in this setting. Accordingly, we adopt a parsimonious specification that focuses on bond-level characteristics, including issue size, time to maturity, and indicators for insured, callable, and general obligation bonds. Standard errors are double-clustered at the bond issue and event level.

3.4 Social media coverage

To capture variation in social media coverage, we focus on Reddit, an online platform organized into topic-specific communities ("subreddits"). In contrast to prior work that studies firm- or asset-specific discussions, we exploit location-based subreddits that provide real-time, ground-level information about local conditions. These communities are typically created and maintained by residents and vary widely in their geographic coverage.

We construct a measure of social media coverage by identifying subreddits corresponding to specific cities or counties based on geographic identifiers in their names and descriptions (e.g.,

[7] Moreover, bonds that do not trade in all four periods of our sample could be considered too thinly traded and stale. As such, the observed yields we use to calculate average yield and the change in yields might be far more impacted by the

r/miami, r/dallas). Each subreddit is mapped to a county using its stated location. We then use the creation date of each subreddit to determine when discussion about a given location first becomes available. A county is classified as covered by social media once at least one associated subreddit exists.[8] Accordingly, our Reddit measure should be interpreted as capturing the presence of local online communities that could facilitate decentralized information sharing, rather than the verified flow of disaster-specific information.

This measure generates a staggered introduction of social media coverage across counties, which provides systematic variation in the information environment. In the context of the framework described in Section 2, the emergence of a local subreddit expands the set of signals available to investors prior to the arrival of official information. We exploit this staggered roll-out by interacting social media coverage with event indicators, allowing us to identify how the availability of additional decentralized signals affects price dynamics. Specifically, we expand the model in Equation (2), which uses asubset of affected counties, by adding interaction with Reddit coverage:[9]

$$\Delta Yield_{i,c,g,t} = \beta_0 + \beta_1 CoveragepGap_{c,g,t} + \beta_2 Reddit\ Coverage_{c,g,t} + \beta_3 CoveragepGap_{c,g,t} \times Reddit\ Coverage_{c,g,t} + Bond\ Controls + Fixed\ Effects + \epsilon_{i,c,g,t} \qquad (3)$$

As shown in Figure 4, the presence of local communities increases steadily over time, generating substantial cross-sectional and temporal variation. This variation allows us to compare otherwise

[8] Consistent with our definition of traditional media coverage, if a subreddit corresponds to a city located within a county, we consider it as if the subreddit is covering the whole county.

[9] We use the Equation (2) model for ease of interpretation, to avoid triple interactions. However, we observe similar patterns when using model from Equation (1) with added *Reddit Coverage.*

similar counties exposed to the same shocks but differing in the availability of social media as an information source.

Insert Figure 4

4. Results

4.1 Baseline Price Dynamics and Information Gaps

We begin by examining how municipal bond yields respond to exogenous shocks in the absence of complete information. Our empirical setting provides a natural test of this mechanism: natural disasters generate immediate shocks to local fundamentals, while official information about their magnitude is revealed with delay. If investors face information frictions, prices should initially reflect uncertainty and be subsequently revised as more precise information becomes available. Consistent with this prediction, we find a clear pattern of initial repricing followed by correction, concentrated in environments with limited information availability.

Insert Table 1

Table 1 reports the evolution of yield changes around disaster events. We report the results using Equation (1) in Panel A, and the results using a restricted sample on only affected counties and the simplified model from Equation (2) in Panel B. Both results are consistent. We observe a sharp increase in yields for bonds in affected counties with *Coverage gap*, i.e., in locations not covered by traditional media, in the initial period following the disaster. This indicates that investors revise upward the perceived risk of affected bonds due to uncertainty about the extent of damage. In the subsequent period, when FEMA begins issuing information, yield changes are small and statistically indistinguishable from zero. However, in the following month, when a more complete set of information should be available from FEMA, yields from locations without media coverage exhibit a significant reversal, partially offsetting the initial increase.

This evidence is consistent with an information-based interpretation. In the absence of coverage, investors observe the occurrence of the shock but lack information about its magnitude, leading to an initial overreaction that is later corrected (Da et al., 2011; Engelberg & Parsons, 2011). When media coverage is present, prices incorporate more information earlier, reducing the need for subsequent adjustment. These findings suggest that gaps in traditional information intermediaries can generate predictable patterns of mispricing following exogenous shocks. We next examine whether alternative information sources can mitigate these information frictions.

4.2 Social media and alleviation of information gaps

We next examine whether decentralized information from social media mitigates the information frictions documented above. As discussed in Section 3, the staggered introduction of location-specific Reddit communities generates systematic variation in the availability of local information across counties. In the framework outlined in Section 2, the presence of social media expands the set of signals available to investors during the period in which official information is absent. If these signals are informative, prices should incorporate more information following the shock, reducing the need for subsequent correction.

Insert Table 2

Table 2 examines whether the yield dynamics documented in the baseline results differ in counties with local Reddit coverage. We observe that yields are significantly impacted only in the first period, i.e., following the disaster. The effect of yield coverage seems to be attenuating the effect in the interaction term. We do not observe any corrections in the subsequent periods. This indicates that prices incorporate information about the magnitude of the shock during the initial post-event period, prior to the release of official estimates.

The contrast with the baseline results is stark. In the absence of social media coverage, yields exhibit a clear pattern of initial reaction followed by subsequent correction, consistent with investors updating their beliefs as more information becomes available. When social media coverage is present, this pattern disappears: prices adjust following the shock without the need for later revision. This difference arises despite the fact that counties are exposed to the same underlying events, suggesting that variation in information availability—rather than differences in fundamentals—drives the observed dynamics.

These findings support an information-based interpretation of social media. The role of social media is not to accelerate price adjustment in a mechanical sense, but to expand the information set available to investors. Even if the signals provided by social media are imperfect, they allow market participants to form more accurate beliefs about the magnitude of the shock, reducing uncertainty and eliminating the need for subsequent correction. In this sense, social media does not simply add noise to the information environment; rather, it provides economically relevant signals that are incorporated into prices when they are most needed.

4.3 Event severity and information as mechanism

We next examine whether the effects of information availability vary with the severity of the underlying shock. In the framework outlined in Section 2, the value of information is increasing in the magnitude of the shock, as larger shocks generate greater uncertainty about fundamentals and therefore a larger scope for mispricing. If our results are driven by information frictions, we should expect the baseline pattern of initial reaction and subsequent correction to be more pronounced in locations more severely affected.

Insert Table 3

Table 3 reports the results when we split the sample based on disaster severity. In counties experiencing more severe shocks, we observe a strong pattern of initial repricing followed by a significant correction in both subsequent periods. In contrast, in less severely affected counties, the correction is substantially smaller and occurs earlier, with yields stabilizing once initial information begins to arrive. This suggests that the extent of mispricing is closely linked to the magnitude of the underlying shock, consistent with investors facing greater uncertainty in more severely affected areas.

Insert Table 4

We next introduce social media coverage into this framework and report the results in Table 4. The results show that the role of social media is strongest precisely in the high-severity subsample. In these counties, social media coverage significantly attenuates the initial increase in yields and eliminates the subsequent correction. In contrast, the effects are limited in low-severity counties, where the baseline mispricing is already modest. This pattern indicates that social media, as alternative information sources, are most valuable when the underlying shock is large and the informational environment is most uncertain.

These findings further support an information-based interpretation of our results. When the magnitude of the shock is large, the cost of remaining uninformed is higher, and additional signals—such as those provided by social media—play a greater role in shaping beliefs. Conversely, when the shock is small, the benefits of additional information are limited, and prices adjust with little need for subsequent revision. The interaction between severity and social media coverage thus provides direct evidence that the effects we document operate through information frictions rather than alternative channels.

Taken together, the results show that the impact of social media is not uniform, but instead depends on the informational needs of the market. Social media is most valuable precisely in those environments where traditional sources provide insufficient guidance and uncertainty about fundamentals is greatest.

5. Conclusion

This paper studies how social media affects price formation in environments characterized by limited information. We exploit a setting with exogenous shocks and a well-defined delay in official reporting, which creates periods in which investors observe the occurrence of an event but lack reliable information about its magnitude. In these environments, we document a systematic pattern of initial repricing followed by subsequent correction, consistent with prices reflecting uncertainty and being revised as information becomes available. This pattern is concentrated in locations without traditional media coverage, highlighting the role of information gaps in generating predictable price dynamics.

We show that social media mitigates these information frictions. Using the staggered introduction of location-based online communities, we identify variation in the availability of decentralized information across otherwise similar counties exposed to the same shocks. In locations with social media coverage, prices incorporate information during the initial post-event period, and the subsequent correction disappears. This effect is strongest when the underlying shock is more severe, consistent with the value of additional information increasing with the magnitude of uncertainty. Taken together, these findings indicate that social media improves market outcomes by expanding the information set available to investors when traditional sources are absent. Furthermore, these findings have implications for a wide range of settings in which information about fundamentals is incomplete or unevenly distributed.

More broadly, our results highlight that market efficiency depends not only on the precision of information, but also on its availability. Social media is often viewed as a source of noise, with contemporary literature focusing on its distortive effects. Our findings show that it can play a central role in facilitating price discovery in settings where conventional information intermediaries provide limited coverage. This perspective suggests that the impact of social media on financial markets cannot be evaluated in isolation, but must be understood in the context of the broader information environment.

References

Addoum, J. M., Ng, D. T., & Ortiz-Bobea, A. (2020). Temperature shocks and establishment sales. *The Review of Financial Studies*, 33(3), 1331-1366.

Barber, B. M., Huang, X., Jorion, P., Odean, T., & Schwarz, C. (2024). A (sub) penny for your thoughts: Tracking retail investor activity in TAQ. *The Journal of Finance*, 79(4), 2403-2427.

Bradley, D., Hanousek Jr, J., Jame, R., & Xiao, Z. (2024). Place your bets? The value of investment research on Reddit's Wallstreetbets. *The Review of Financial Studies*, 37(5), 1409-1459.

Bradley, D., Hanousek Jr., J., Hanousek, J., & Svoboda, D. (2026). Bot-Induced Social Media Manipulation and Stock Market Distortion. Available at SSRN 5912222.

Bushman, R., & Pinto, J., (2023). The influence of short selling on negative press coverage of firms. *Management Science*, 70 (3), 1924–1942.

Cookson, J. A., & Niessner, M. (2020). Why don't we agree? Evidence from a social network of investors. *The Journal of Finance*, 75(1), 173-228.

Cookson, J. A., Engelberg, J. E., & Mullins, W. (2023). Echo chambers. *The Review of Financial Studies*, 36(2), 450-500.

Cookson, J.A., Lu, R., Mullins, W., & Niessner, M., (2024). The social signal. *Journal of Financial Economics*, 158 (8), 103870.

Da, Z., Engelberg, J., & Gao, P. (2011). In search of attention. *The journal of finance*, *66*(5), 1461-1499.

Diamond, D. W., & Verrecchia, R. E. (1981). Information aggregation in a noisy rational expectations economy. *Journal of Financial Economics*, 9(3), 221-235.

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

Fama, E. F. (1970). Efficient capital markets: A review of theory and empirical work. *The journal of Finance*, 25(2), 383-417.

Gao, P., Lee, C., & Murphy, D. (2020). Financing dies in darkness? The impact of newspaper closures on public finance. *Journal of Financial Economics*, 135(2), 445-467.

Goldstein, I., Ozdenoren, E., & Yuan, K. (2013). Trading frenzies and their impact on real investment. *Journal of Financial Economics*, 109(2), 566-582.

Goldstein, I., Xiong, Y., & Yang, L. (2025). Information sharing in financial markets. *Journal of Financial Economics*, 163, 103967.

Green, R. C., Hollifield, B., & Schürhoff, N. (2007). Dealer intermediation and price behavior in the aftermarket for new bond issues. *Journal of Financial Economics*, 86(3), 643-682.

Grossman, S. J., & Stiglitz, J. E. (1980). On the impossibility of informationally efficient markets. *The American Economic Review*, 70(3), 393-408.

Gurun, U. G., & Butler, A. W. (2012). Don't believe the hype: Local media slant, local advertising, and firm value. *The Journal of Finance*, 67(2), 561-598.

Hanousek Jr, J., Hanousek, J., & Sokolov, K. (2026). 𝕏 Bots and Earnings Announcements. Available at SSRN 4595101.

Harris, L. E., & Piwowar, M. S. (2006). Secondary trading costs in the municipal bond market. *The Journal of Finance*, 61(3), 1361-1397.

Hung, M., Li, X., & Wang, S. (2015). Post-earnings-announcement drift in global markets: Evidence from an information shock. *The Review of Financial Studies*, 28(4), 1242-1283.

Mola, S., Rau, P. R., & Khorana, A. (2013). Is there life after the complete loss of analyst coverage?. *The Accounting Review*, 88(2), 667-705.

Morris, S., & Shin, H. S. (2002). Social value of public information. *The American Economic Review*, 92(5), 1521-1534.

Painter, M. (2020). An inconvenient cost: The effects of climate change on municipal bonds. *Journal of Financial Economics*, 135(2), 468-482.

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.

Vives, X. (2008). Information and learning in markets: the impact of market microstructure. Princeton University Press.

You, J., Zhang, B., & Zhang, L. (2018). Who captures the power of the pen? *The Review of Financial Studies*, 31 (1), 43-96.

Figure 1. Information structure and discrete-time yield framework

This figure illustrates the information environment underlying the empirical design. Panel A maps the event into 30-day intervals. The 30-day interval was chosen because FEMA begins reporting damage estimates thirty days after disaster has been declared. Natural disaster occurs is declared in day 0. During the subsequent 30 days, during event time t=0, there are no official damage estimates, with media providing incomplete information for only a select number of locations. Official reporting by FEMA begins in period t=1, and by t=2, information about the magnitude and long term economic impacts of the shock are more complete. While traditional media coverage during this period is selective, social media may provide additional local signals. In Panel B, we show the predicted path of monthly yield changes, defined as $\Delta Yield_t = Yield_t - Yield_{t-1}$. In counties where social media is unavailable in period t, yields increase sharply in time t=0 and there is subsequent correction in t=2, as more complete information becomes available and uncertainty is mostly resolved. In counties where social media coverage is available in period t, the initial pricing of the effect incorporates more information, so the later correction is attenuated or absent.

Panel A

Event time	$t = -1$	$t = 0$	$t = 1$	$t = 2$
	[-30,0)	[0,30)	[30,60)	[60,90]
Information set	Pre-shock baseline	Disaster is declared at time 0. No official damage estimates Traditional media is selective and sparse Social media may provide local signals	FEMA begins reporting estimates Information starts to arrive	More complete public information available from official sources Uncertainty largely resolved

Panel B

	$\Delta Yield_{t=0}$	$\Delta Yield_{t=1}$	$\Delta Yield_{t=2}$
No social media	+	0	–
With social media	+	0	0 / attenuated

Figure 2. Number of disaster events per year

This graph displays the number of disaster events and the share of covered and uncovered events during the coverage window t=0 for the period 2008–2024.

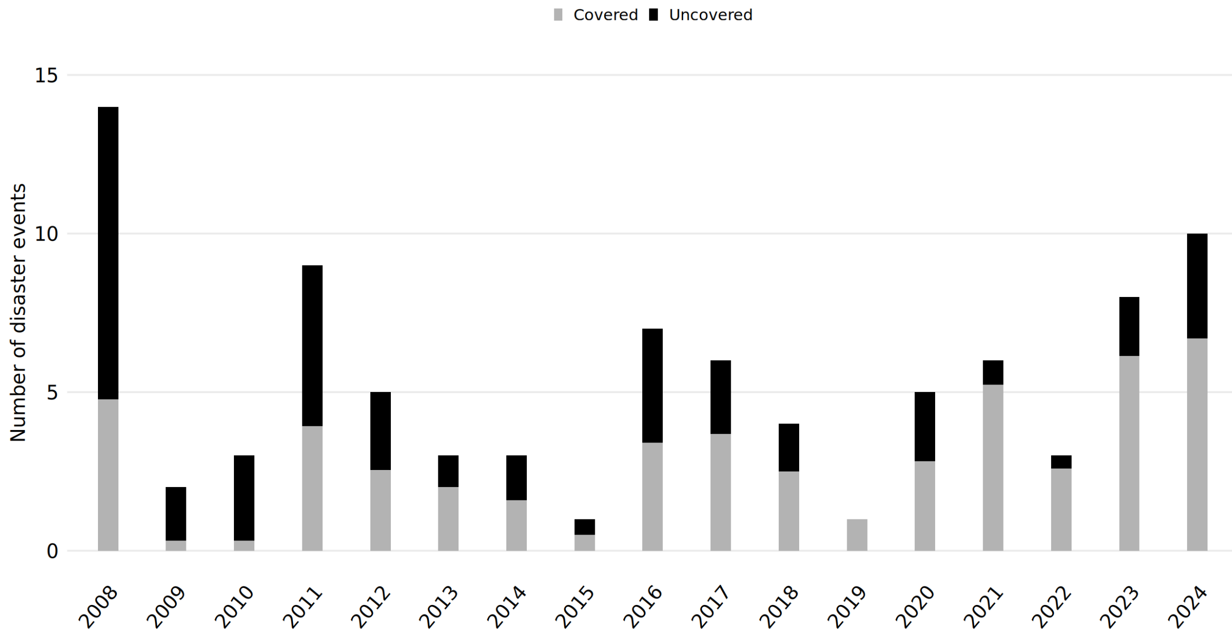

Figure 3. Disaster coverage in the United States

This map identifies counties affected by disaster events and the share of events explicitly mentioned in news coverage during the coverage window t=0. Darker counties indicate locations with a lower share of covered events, while light gray represents locations with a higher share of covered events. Counties shown in white were never affected during the sample period.

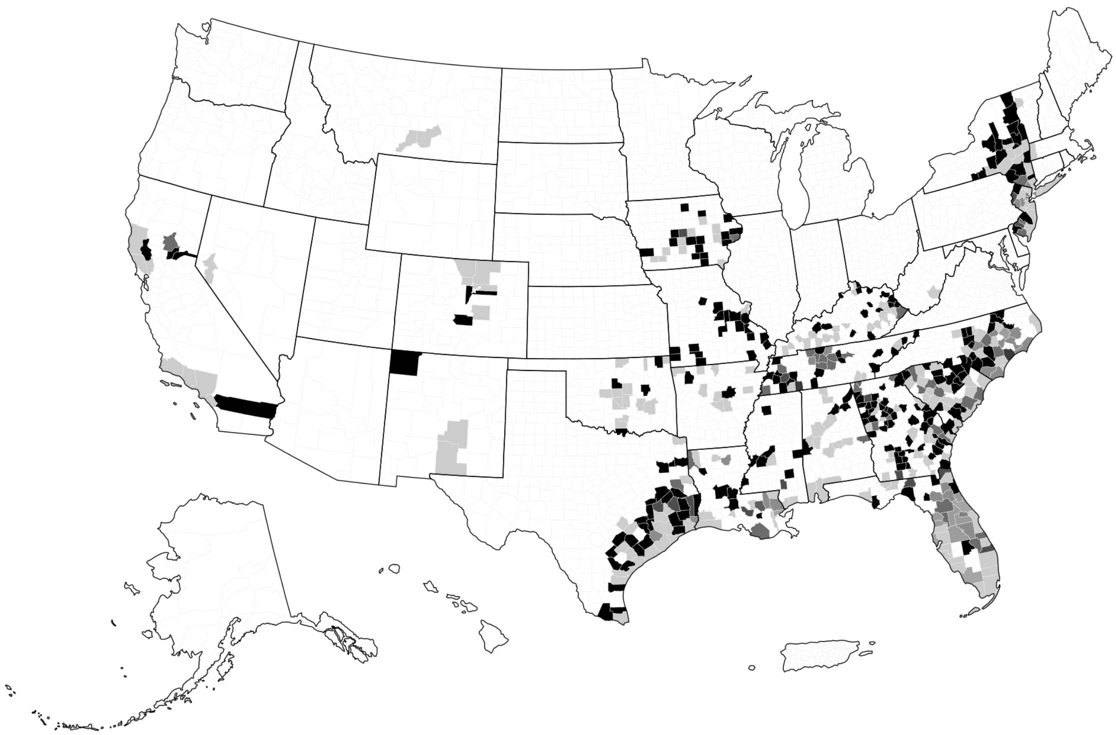

Figure 4. Number of counties with Reddit coverage

This figure displays the number of counties with Reddit coverage per year over the period 2008–2024.

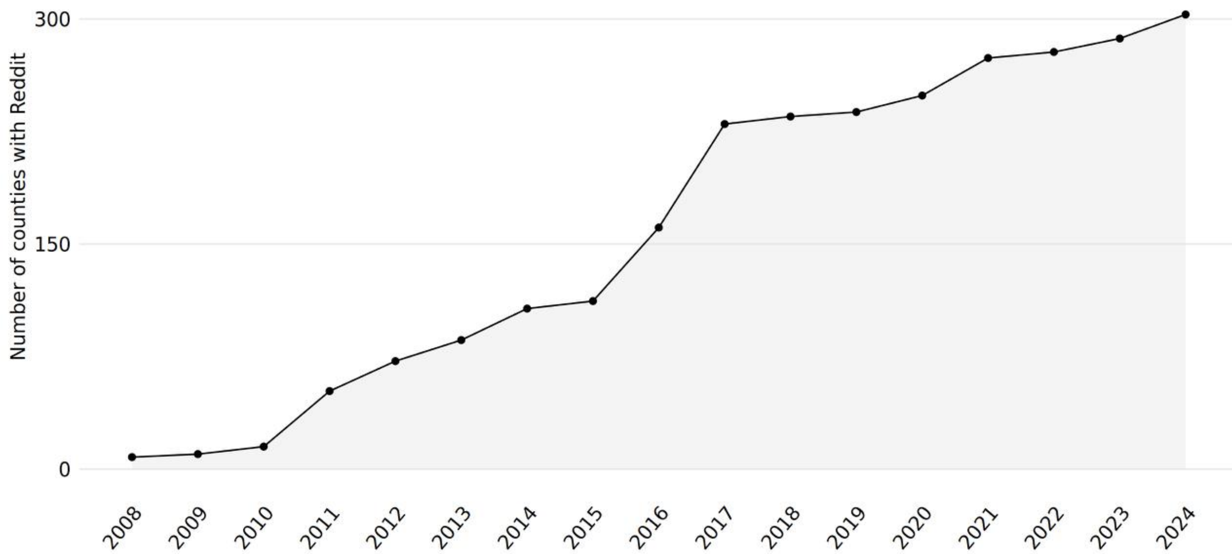

Table 1 – Effect of media silence on price efficiency

This table analyzes the impact of media silence on the change in yields following a disaster. The dependent variable is the change in municipal bond yield measured over consecutive 30-day windows around the disaster declaration. Our key variable of interest is the variable *Coverage Gap*, which is defined as being equal to one if during the 30-days following a disaster, when FEMA is not issuing offical information on the extent of damages, the given location has not been covered by traditional media sources. We define *Affected* as being equal to one if the location was impacted by the event. Our sample for each event only consists of counties in the state(s) impacted by the event. Column (1) captures the yield change between the pre-disaster window and the first 30-day window following the disaster declaration, during the period of the informational black out. Columns (2) and (3) report subsequent 30-day changes. All specifications include county fixed effects, event fixed effects, and rating controls using the lowest available credit rating across Standard & Poor's and Moody's. Due to the extensive fixed effects and the use of differences in yield, we adopt more parsimonious set of control variables: issue size (log), time to maturity (level and square), and indicators for insured, callable, and general obligation bonds. Panel A shows the results using our Equation (1) as the regression specification. Panel B shows the results when we restrict our sample to affected counties and use the simplified regression model specified in Equation (2). The sample is restricted to bonds with observed trading activity in every event-time window (t to t+2) to mitigate effects from thinly traded bonds. All yield change variables are winsorized at the 1st and 99th percentiles. In both panels, standard errors are double-clustered by issue and event-time. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

Panel A)

	ΔY (*t, t-1*)	ΔY (*t+1, t*)	ΔY (*t+2, t+1*)
	(1)	(2)	(3)
Coverage Gap	0.005	0.012	0.003
	(0.007)	(0.009)	(0.006)
Affected	-0.004	-0.012	0.003
	(0.011)	(0.009)	(0.007)
Coverage Gap × Affected	0.030**	-0.007	-0.042***
	(0.014)	(0.018)	(0.011)
County FE	Yes	Yes	Yes
Event FE	Yes	Yes	Yes
Rating controls	Yes	Yes	Yes
Bond controls	Yes	Yes	Yes
N (observations)	28,385	28,259	23,270
R^2	0.36286	0.41003	0.32743
R^2 (adjusted)	0.34852	0.39664	0.30990

Panel B)

	ΔY ($t, t-1$)	ΔY ($t+1, t$)	ΔY ($t+2, t+1$)
	(1)	(2)	(3)
Coverage Gap	0.054***	-0.011	-0.049**
	(0.018)	(0.029)	(0.024)
County FE	Yes	Yes	Yes
Event FE	Yes	Yes	Yes
Rating controls	Yes	Yes	Yes
Bond controls	Yes	Yes	Yes
N (observations)	10,209	10,144	8,496
R^2	0.36666	0.38570	0.23740
R^2 (adjusted)	0.34695	0.36640	0.20987

Table 2 – Attenuating effect of social media

This table analyzes whether social media attenuates the impact of traditional media silence on the change in yields following a disaster. The dependent variable is the change in municipal bond yield measured over consecutive 30-day windows around the disaster declaration. Our key variable of interest is the variable *Reddit Coverage*, which is defined as one if the location has their own subreddit community covering it. The staggered roll-out of Reddit coverage is observed in Figure 4. We further again define *Coverage Gap*, which is defined as being equal to one if during the 30-days following a disaster, when FEMA is not issuing offical information on the extent of damages, the given location has not been covered by traditional media sources. Our sample for each event only consists of impacted counties in the state(s) impacted by the event. Column (1) captures the yield change between the pre-disaster window and the first 30-day window following the disaster declaration, during the period of the informational black out. Columns (2) and (3) report subsequent 30-day changes. All specifications include county fixed effects, event fixed effects, and rating controls using the lowest available credit rating across Standard & Poor's and Moody's. Due to the extensive fixed effects and the use of differences in yield, we adopt more parsimonious set of control variables: issue size (log), time to maturity (level and square), and indicators for insured, callable, and general obligation bonds. The sample is restricted to bonds with observed trading activity in every event-time window (t to t+2) to mitigate effects from thinly traded bonds. All yield change variables are winsorized at the 1st and 99th percentiles. In both panels, standard errors are double-clustered by issue and event. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

	ΔY (*t, t-1*)	ΔY (*t+1, t*)	ΔY (*t+2, t+1*)
	(1)	(2)	(3)
Coverage Gap	0.007	-0.003	-0.038
	(0.025)	(0.033)	(0.025)
Reddit Coverage	-0.067***	0.017	-0.025
	(0.025)	(0.018)	(0.021)
Coverage Gap × Reddit Coverage	0.100***	-0.015	-0.051
	(0.034)	(0.027)	(0.035)
County FE	Yes	Yes	Yes
Event FE	Yes	Yes	Yes
Rating controls	Yes	Yes	Yes
Bond controls	Yes	Yes	Yes
N (observations)	10,209	10,144	8,496
R^2	0.36720	0.38572	0.23773
R^2 (adjusted)	0.34738	0.36629	0.21002

Table 3 – Media silence and event severity

This table analyzes the impact of media silence on the change in yields following a disaster when distinguishing events based on severity. The dependent variable is the change in municipal bond yield measured over consecutive 30-day windows around the disaster declaration. Our key variable of interest is the variable *Coverage Gap*, which is defined as being equal to one if during the 30-days following a disaster, when FEMA is not issuing offical information on the extent of damages, the given location has not been covered by traditional media sources. Our sample for each event only consists of impacted counties in the state(s) impacted by the event. For every yield, we report results separately for high-severity (low-severity) events, defined as in the top (bottom) 40% of the damage distribution based on total FEMA reported damages at the county level. Columns (1) and (2) capture the yield change between the pre-disaster window and the first 30-day window following the disaster declaration, during the period of the informational black out. Columns (3)-(4) and (5)-(6) report subsequent 30-day changes. All specifications include county fixed effects, event fixed effects, and rating controls using the lowest available credit rating across Standard & Poor's and Moody's. Due to the extensive fixed effects and the use of differences in yield, we adopt more parsimonious set of control variables: issue size (log), time to maturity (level and square), and indicators for insured, callable, and general obligation bonds. The sample is restricted to bonds with observed trading activity in every event-time window (t to t+2) to mitigate effects from thinly traded bonds. All yield change variables are winsorized at the 1st and 99th percentiles. In both panels, standard errors are double-clustered by issue and event. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

	ΔY (*t, t-1*)		ΔY (*t+1, t*)		ΔY (*t+2, t+1*)	
	High-severity	Low-severity	High-severity	Low-severity	High-severity	Low-severity
	(1)	(2)	(3)	(4)	(5)	(6)
Coverage Gap	0.047***	-0.108***	-0.031***	0.074**	-0.038**	0.055
	(0.011)	(0.037)	(0.010)	(0.036)	(0.018)	(0.044)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Event FE	Yes	Yes	Yes	Yes	Yes	Yes
Rating controls	Yes	Yes	Yes	Yes	Yes	Yes
Bond controls	Yes	Yes	Yes	Yes	Yes	Yes
N (observations)	8,557	7,737	9,164	7,859	5,616	4,534
R^2	0.35847	0.27068	0.18819	0.43112	0.30780	0.16943
R^2 (adjusted)	0.34178	0.24602	0.16886	0.41250	0.28514	0.13209

Table 4 – Social media coverage and event severity

The dependent variable is the change in municipal bond yield measured over consecutive 30-day windows around the disaster declaration. Our key variable of interest is the variable *Reddit Coverage*, which is defined as one if the location has their own subreddit community covering it. The staggered roll-out of Reddit coverage is observed in Figure 4. We further again define *Coverage Gap*, which is defined as being equal to one if during the 30-days following a disaster, when FEMA is not issuing offical information on the extent of damages, the given location has not been covered by traditional media sources. Our sample for each event only consists of impacted counties in the state(s) impacted by the event. For every yield, we report results separately for high-severity (low-severity) events, defined as in the top (bottom) 40% of the damage distribution based on total FEMA reported damages at the county level. Columns (1) and (2) capture the yield change between the pre-disaster window and the first 30-day window following the disaster declaration, during the period of the informational black out. Columns (3)-(4) and (5)-(6) report subsequent 30-day changes. All specifications include county fixed effects, event fixed effects, and rating controls using the lowest available credit rating across Standard & Poor's and Moody's. Due to the extensive fixed effects and the use of differences in yield, we adopt more parsimonious set of control variables: issue size (log), time to maturity (level and square), and indicators for insured, callable, and general obligation bonds. The sample is restricted to bonds with observed trading activity in every event-time window (t to t+2) to mitigate effects from thinly traded bonds. All yield change variables are winsorized at the 1st and 99th percentiles. In both panels, standard errors are double-clustered by issue and event. ***, **, and * denote statistical significance at 1%, 5%, and 10%, respectively.

	ΔY (*t, t-1*)		ΔY (*t+1, t*)		ΔY (*t+2, t+1*)	
	High-severity	Low-severity	High-severity	Low-severity	High-severity	Low-severity
	(1)	(2)	(3)	(4)	(5)	(6)
Coverage Gap	0.012	-0.126***	-0.019	0.056***	-0.038	0.024
	(0.015)	(0.042)	(0.018)	(0.018)	(0.025)	(0.066)
Reddit	-0.038**	-0.010	-0.019	0.004	-0.014	-0.018
	(0.015)	(0.022)	(0.019)	(0.014)	(0.020)	(0.027)
Coverage Gap × Reddit	0.096***	0.064	-0.045	0.072	-0.006	0.093
	(0.015)	(0.069)	(0.027)	(0.056)	(0.038)	(0.055)
County FE	Yes	Yes	Yes	Yes	Yes	Yes
Event FE	Yes	Yes	Yes	Yes	Yes	Yes
Rating controls	Yes	Yes	Yes	Yes	Yes	Yes
Bond controls	Yes	Yes	Yes	Yes	Yes	Yes
N (observations)	8,557	7,737	9,164	7,859	5,616	4,534
R^2	0.35887	0.27072	0.18838	0.43116	0.30783	0.16951
R^2 (adjusted)	0.34203	0.24587	0.16888	0.41239	0.28490	0.13178