



DEPARTMENT OF ECONOMICS
DISCUSSION PAPER SERIES

Social media, news media and the stock market

Peiran Jiao, Andre Veiga, Ansgar Walther

Number 805
October, 2016

Manor Road Building, Oxford OX1 3UQ

Social media, news media and the stock market^{*}

Peiran Jiao Andre Veiga Ansgar Walther[†]

This version: July 6, 2016

Abstract

We contrast the impact of traditional news media and social media coverage on stock market volatility and trading volume. We develop a theoretical model of asset pricing and information processing, which allows for both rational traders and a variety of commonly studied behavioral biases. The model yields several novel and testable predictions about the impact of news and social media on asset prices. We then test the model's theoretical predictions using a unique dataset which measures coverage of individual stocks in social and news media using a broad spectrum of print and online sources. Stocks with high social media coverage in one month experience high idiosyncratic volatility of returns and trading volume in the following month. Conversely, stocks with high news media coverage experience low volatility and low trading volume in the following month. These effects are statistically and economically significant and robust to controlling for stock and time fixed effects, as well as time-varying stock characteristics. The empirical evidence on news media is consistent with a market in which some traders are overconfident when interpreting new information. The evidence on social media is consistent with Tetlock (2011)'s "stale news" hypothesis (investors treat repeated information on social networks as though it were new) and with a model where investors' perceptions are subject to random sentiment shocks.

JEL Classification Codes: G02, G12, G14.

Keywords: Social media, news media, behavioral finance, volatility, trading volume.

^{*}We are grateful to Basile Grassi, Ryoko Ito, Dan Quigley, Tarun Ramadorai, Selma Walther and Mungo Wilson for their comments.

[†]Jiao and Veiga: Department of Economics and Nuffield College, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK. Walther: Department of Economics, Nuffield College, and INET@OMS, University of Oxford, Manor Road Building, Oxford OX1 3UQ, UK. Email: peiran.jiao@economics.ox.ac.uk; andre.veiga@economics.ox.ac.uk; ansgar.walther@economics.ox.ac.uk.

1 Introduction

A large academic literature has studied the impact of news on stock markets, and its results suggest that market reactions often deviate from the prescriptions of classical financial theories.¹ However, the existing literature has not significantly addressed the recent rise of social media as a provider of financial news.² In this paper, we show that stock markets respond very differently to news media and social media. We develop a theoretical model of asset pricing and information processing, which allows for a variety of commonly studied forms of bounded rationality and yields several novel and testable predictions. We use these predictions to determine which forms of bounded rationality are compatible with observed market outcomes using a unique new dataset. Our results suggest that some traders neglect the correlation between signals from social media sources (“stale news”), or that traders’ perception of social media signals is subject to volatile sentiments. On the other hand, when interpreting signals from traditional news media, our results suggest that traders are subject to overconfidence.

To guide our analysis, we begin by extending a standard CARA-Gaussian asset pricing model to allow for full rationality, but also for several deviations from rationality present in the financial and psychological literature (Barberis and Thaler, 2003). Namely, our model can accommodate (i) *Overconfident* investors who overestimate the precision of their information as in Scheinkman and Xiong (2003); (ii) investors who treat *stale news* as though it were fresh information as in Tetlock (2011); (iii) *conservative* investors who are wedded to their prior beliefs; (iv) investors who cannot process all available information due to *attention constraints* as in Peng and Xiong (2006); and (v) investors whose perception of information is distorted by *sentiments* or *confirmation bias* as in Rabin and Schrag (1999). Each of these specifications yields distinct testable predictions about the effect of additional coverage (by news or social media) on market outcomes.

We proceed to test the model’s predictions on a novel dataset that describes the coverage of individual stocks in social and news media, obtained from the Thomson Reuters MarketPsych Indices (TRMI) database. These measures are aggregated from a broad spectrum of news media sources, both in print and online, and the most popular social media content. We merge this dataset with stock-level prices and characteristics at a monthly frequency. A

¹For instance, investors are prone to overreact to new information (De Bondt and Thaler, 1985), to react to “stale news” which merely repeat previous revelations (Tetlock, 2011), and to focus on “attention-grabbing” stocks rather than considering all available information (Barber and Odean, 2008).

²Facebook and Twitter, founded just over ten years ago, now have around 1.8 billion active users between them, and some 500 million tweets are sent daily. A significant share of social media activity is related to stock markets, and indeed a number of start-up firms now use algorithms to extract market-relevant information from social media. See, for example: <http://www.wsj.com/articles/tweets-give-birds-eye-view-of-stocks-1436128047>

key quantity of interest is the “buzz” around a stock, defined as the number of stock-specific phrases or words used to calculate TRMI in news articles or social media posts, relative to aggregate coverage of the stock market. We investigate the effect of social media and news media buzz on two key measures of market activity, namely idiosyncratic return volatility and trading volume.

We find that stocks with large *social media* buzz have significantly *higher* idiosyncratic volatility and *higher* trading volume over the next month. Relative to a stock with no social media buzz, the most talked-about stocks on social media experience an increase in their average volatility of about 50%, and an increase in average trading volume of 25%, over the next month. This empirical evidence is not consistent with a model where signals are processed in a fully rational way. Instead, it is consistent with a model of “stale news” where social media buzz contains, at least in part, repetitions of existing information, but where some investors treat these repetitions as fresh information. Thus, our results are in line with the Tetlock’s (2011) hypothesis that “stale news” affects markets, and further suggest that these effects are particularly significant for the case of social media. The empirical results on social media are also consistent with a model of “strong sentiments,” where investors’ interpretation of information is subject to volatile and correlated biases.

Conversely, we find that stocks with large *news media* buzz have significantly *lower* idiosyncratic volatility and *lower* trading volume over the next month. The stocks most heavily covered in news media experience a decrease of about 75% in average volatility and 40% in average trading volume. Following our theoretical predictions, this empirical evidence is consistent with a model in which some investors are overconfident, and believe that the precision of the news media signals they observe is higher than their true precision. In this case, observing a large sample of information reduces the likelihood of large disagreements between traders and therefore dampens market activity which is driven by overconfidence.

Therefore, our empirical results strongly reject the hypothesis that the processing of information from social and news media is described by the same model, since the two consistently have opposite effects on stock market volatility and trading volume. Both findings are robust to controlling for stock and time fixed effects, as well as for time-varying stock characteristics. The effects are also economically significant.

A potential concern about our reading of the evidence is that social media buzz is endogenous to market activity. For instance, an exogenous shock may increase both buzz and volatility; an increase in volatility might increase the demand for news, make traders more willing to participate in social media, and thereby increase buzz; alternatively, traders might post their opinions on social media after placing orders so as to “talk their own book.” However, if any of these were true, we would expect endogenous buzz and market activity to

increase simultaneously. By contrast, our empirical results indicate that *current* buzz predicts *future* market activity even when we control for *current* market activity. Therefore, it is unlikely that our results can be fully explained by the presence of endogeneity.

This paper relates to the literature on news, communication and stock markets. Regarding news coverage, [Cutler et al. \(1989\)](#) point out that macroeconomic and political news do not explain a large portion of stock market movements at the monthly frequency. Using more granular news data from the *Wall Street Journal* and shorter time intervals, [Tetlock \(2007\)](#) demonstrates that textual measures of news sentiment can be used to predict aggregate market returns a few days ahead, and that a high absolute value of sentiment in the news media (i.e. either severe optimism or severe pessimism) predicts high trading volume the next day. [Fang and Peress \(2009\)](#) show that stocks with low media coverage have higher average returns, especially if they have high volatility and low institutional ownership or analyst following. [Bushee et al. \(2010\)](#) show that greater press coverage reduces information asymmetries and is associated with larger reactions of prices and trading volumes to earnings announcements.

Regarding communication, [Coval and Shumway \(2001\)](#) show that ambient noise levels in a trading pit predict high intraday volatility and low market depth. [Antweiler and Frank \(2004\)](#), in what is perhaps the most closely related paper to ours, show that the number of messages posted in specialist chat rooms such as *RagingBull* positively predicts volatility and trading volume over the next day. [Chen et al. \(2014\)](#) construct a measure of negative and positive words on *SeekingAlpha*, which can predict abnormal returns over a three-month horizon. [Da et al. \(2015\)](#) construct a sentiment indicator using *Google Trends*, which can predict return reversals and has a temporary (two-day) effect on volatility.

Our contribution in this context is twofold: First, we *directly* contrast the effects of two distinct news sources (social and news media) and point out the stark differences between the two. This comparison is facilitated by our unique dataset, which also has the advantage of drawing from a more comprehensive set of media sources than existing studies. Second, we relate our empirical results to theoretical predictions, and demonstrate which theoretical models of boundedly rational information processing are consistent with the empirical evidence.

Another contribution, therefore, is to the behavioral asset pricing literature, which [Hirshleifer \(2001\)](#) and [Barberis and Thaler \(2003\)](#) review in detail. Existing theoretical papers analyze the behavior of markets where investors are overconfident ([Kyle and Wang, 1997](#); [Odean, 1998](#); [Daniel et al., 1998](#); [Gervais and Odean, 2001](#); [Scheinkman and Xiong, 2003](#)), conservative ([Barberis et al., 1998](#)) and subject to confirmation bias ([Rabin and Schrag, 1999](#); [Pouget et al., 2014](#)). Our model of information processing offers a parsimonious way of

determining the empirical predictions of several behavioral biases. Moreover, we place more emphasis on investor heterogeneity than the existing literature, which allows us to derive new predictions about trading volume and return volatility.

The remainder of the paper is structured as follows. In Section 2, we set up our theoretical framework, and the main empirical implications are derived in Section 2.5. In Section 3, we describe our data on news media, social media and financial variables, and present summary statistics. In Section 4, we show and interpret our main empirical results and check their robustness. Section 5 concludes.

2 Model

We study a model of a stock market with two dates $t \in \{0, 1\}$. At date 0, a unit mass of investors trade a risk-free asset with return r and a risky asset (stock) with market price p . The stock is in zero net supply and yields a random payoff θ per unit at date 1. All investors have constant absolute risk aversion (CARA) with parameter γ .

Our aim is to derive empirical implications by characterizing how high media coverage or “buzz” at date 0 affects the observed volatility of stock returns and the trading volume of stocks traded between dates 0 and 1. We allow for both rational traders and behavioral traders. The latter process information subject to a variety of biases that have been identified by the behavioral finance literature. We begin by proposing a general model of rational and behavioral traders, and applying it to the most commonly studied deviations from rationality. In Section 2.5, we summarize our empirical predictions.

2.1 Beliefs and biases

Investors have a common prior belief that the stock’s payoff θ is normally distributed: $\theta \sim \mathcal{N}(\theta_0, \rho_0^{-1})$. At date 0, they observe a vector of N noisy public signals $\mathbf{s} = (s_1, \dots, s_i, \dots, s_N)$, where $s_i = \theta + \epsilon_i$ and the variables $\epsilon_i \sim \mathcal{N}(\theta_0, \rho_i^{-1})$ are i.i.d. The parameter ρ_i measures the precision of the i^{th} signal. We interpret the number of signals N as the “buzz” in either social or news media.³

A mass $1 - \lambda$ of investors are *rational* and update beliefs in line with Bayes’ rule. Thus,

³In our empirical analysis, we focus on a normalized measure, *relative buzz*, which is the ratio of the number of times an individual stock is mentioned to the number of articles about the stock market as a whole. In the model, we interpret increased buzz as an increase in the *total* number of signals about the stock in question. However, the predictions of this model are relevant for our empirical measure because, holding constant the amount of buzz around other assets, an increase in the number of signals about an individual stock corresponds to an increase in relative buzz also.

after observing signals \mathbf{s} they believe that $\theta|\mathbf{s} \sim \mathcal{N}(\theta_R, \rho_R^{-1})$, where

$$\theta_R = \sum_i w_i s_i + \left(1 - \sum_i w_i\right) \theta_0, \quad (1)$$

$$\rho_R = \rho_0 + \sum_i \rho_i. \quad (2)$$

The posterior mean is a weighted average of prior beliefs and average signals, where $w_i = \rho_i/\rho_R$ is the weight attached to information with precision ρ_i . The posterior precision increases linearly with the aggregate precision of signals ($\sum_i \rho_i$).

A mass λ of investor are *behavioral* and their processing of information deviates from Bayes' rule. However, we assume that their beliefs preserve two key features: posterior beliefs inherit the normality of priors, and are linear in signals. Accordingly, behavioral investors believe that $\theta|\mathbf{s} \sim \mathcal{N}(\theta_B, \rho_B^{-1})$, where

$$\theta_B = \sum_i \hat{w}_i (s_i + x_i) + \left(1 - \sum_i \hat{w}_i\right) \theta_0, \quad (3)$$

$$\rho_B = \hat{\rho}_0 + \sum_i \hat{\rho}_i. \quad (4)$$

The updating rules of behavioral traders shown above exhibit three deviations from Bayes' rule. First, the behavioral posterior mean is again a weighted average, but the weight attributed to signals by behavioral traders (\hat{w}_i) may differ from the rational weights w_i . Second, the precision attributed to priors ($\hat{\rho}_0$) and signals ($\hat{\rho}_i$) when deriving the posterior precision can differ from the true ρ_0 and ρ_i . Finally, the perception of the levels of the signals can differ from the truth by a (potentially stochastic) term x_i .

This specification, while requiring several assumptions, has the advantage of being able to capture some of the most common biases studied in behavioral finance.⁴ We now propose a catalog of biases that can be captured by this model, which we use to derive testable predictions in Section 2.5.

Example 1: Overconfidence. A well-documented deviation from Bayesian signal processing is a tendency to overestimate the precision of one's knowledge.⁵ This is commonly modeled by assuming that investors behave like Bayesians but overstate the precision of the signals they receive (Kyle and Wang, 1997; Odean, 1998; Daniel et al., 1998; Gervais and

⁴Barberis and Thaler (2003) and Hirshleifer (2001) give comprehensive surveys of the financial and psychological literature based on these and other biases.

⁵Early experimental evidence was presented by Fischhoff et al. (1977) and Alpert and Raiffa (1982), for example. Odean (1998) reviews the large empirical literature demonstrating overconfidence. Overconfidence is also related to *base rate under-weighting*, where individuals pay too little attention to prior probabilities when updating beliefs (Kahneman and Tversky, 1973).

Odean, 2001; Scheinkman and Xiong, 2003). In our model, assume that investors perceive the correct signals ($x_i = 0$) but believe their precision to be $\hat{\rho}_i = (1 + a)\rho_i$, where $a > 0$ measures overconfidence. Suppose, for simplicity, that all signals were equally precise ($\rho_i = \rho_\epsilon, \forall i$). Then, rational traders would use the Bayesian weights $w_i = \rho_\epsilon/(\rho_0 + N\rho_\epsilon) \equiv w, \forall i$, while behavioral traders use the overconfident weights

$$\hat{w}_i = \frac{\rho_\epsilon(1 + a)}{\rho_0 + N\rho_\epsilon(1 + a)} \equiv \hat{w} > w, \forall i. \quad (5)$$

Example 2: Stale news. Tetlock (2011) shows that news articles have an impact on trading behavior and asset returns, even when the articles in question are “stale news” in the sense that they repeat information that was already available. We can model stale news effects by assuming that there are M truly informative signals (s_1, s_2, \dots, s_M) with precision $\rho_i = \rho_\epsilon$. Then, each of these signals is repeated K times, which gives a total of $N \equiv MK$ signals. The repetitions create a vector of “stale news” signals ($s_{M+1}, s_{M+2}, \dots, s_{MK}$) whose effective precision is zero: $\rho_i = 0$ for $i > M$.⁶

Rational traders recognize that repeated signals are stale news and attach zero weight to them. The rational weights for updating in equation 1 are thus $w_i = \rho_\epsilon/(\rho_0 + M\rho_\epsilon)$ for $i \leq M$ and $w_i = 0$ for $i > M$. By contrast, behavioral traders interpret stale news as new signals with precision $\hat{\rho}_i = \rho_\epsilon$.⁷ Consequently, they attach weight $\hat{w}_i = \rho_\epsilon/(\rho_0 + MK\rho_\epsilon)$ to all signals.

Example 3: Conservatism. The flip-side of overconfidence is a tendency to attach too much weight to one’s prior beliefs (Edwards, 1968). Conservatism is particularly common when individuals have information from large samples, but continue to place higher weight on their prior beliefs than Bayes’ rule would suggest (Griffin and Tversky, 1992). We can capture conservatism by assuming that behavioral investors perceive the correct signals ($x_i = 0$), but believe the precision of their prior to be $\hat{\rho}_0 = (1 + b)\rho_0$ for $b > 0$.⁸ Then, rational traders use the same weights as in Example 1, while behavioral traders use the conservative weights

$$\hat{w}_i = \frac{\rho_\epsilon}{\rho_0(1 + b) + N\rho_\epsilon} \equiv \hat{w} < w, \text{ for all } i. \quad (6)$$

⁶This argument is somewhat informal, because we previously assumed that the errors ϵ_i in different signals are uncorrelated. Modeling stale news more rigorously requires correlated errors – indeed, the errors in a signal and its repeated version are perfectly correlated, and the mistake of behavioral investors is to ignore this correlation. To save notation, we assume instead that the effective precision of the repeated signals is zero, which yields equivalent results.

⁷This bias – the failure to recognize when information is repeated – is related to “persuasion bias” as modeled by DeMarzo et al. (2003).

⁸Barberis et al. (1998) model conservatism by assuming that an investor believes earnings shocks to be persistent when true earnings are a random walk. This is qualitatively similar to our specification: The biased perception of persistence causes their investor to put insufficient weight on earnings announcements.

Example 4: Inattention. Cognitive constraints imply that market participants, especially when acting as individual investors, cannot process all available relevant information. Sims (2003) models cognitive constraints by assuming that agents can only observe signals whose Shannon information lies below a certain bound. The Shannon information of a vector of signals is proportional to the average reduction in entropy – a measure of uncertainty about the fundamental state of the world – that is brought about by observing the signals. In a model with Gaussian signals and priors, the reduction in entropy from observing k signals with precision ρ_ϵ is

$$\frac{1}{2} \log_2 \left(1 + \frac{k\rho_\epsilon}{\rho_0} \right) \equiv I(k).$$

We can introduce inattention in our model by assuming that all signals have precision $\rho_i = \rho_\epsilon$, and that behavioral traders choose to observe a subset of k signals, subject to the constraint that $I(k) \leq \kappa$ for some $\kappa > 0$.⁹ When the total number of signals N is large enough, or when cognitive capacity κ is low enough, then behavioral investors cannot process all relevant signals and $K < N$, while rational traders process all N signals. To ensure that attention is a binding constraint, we will assume that behavioral traders cannot learn from prices in order to infer the information processed by their rational peers.¹⁰

Example 5: Sentiments and confirmation bias. The term x_i in equation (3), which skews the perception of signals by behavioral investors, can be interpreted as a measure of sentiments or market moods (Baker and Wurgler, 2007). For instance, if $E[x_i] > 0$, then investors are generally optimistic and, on average, they perceive signals through “rose tinted glasses” which make all news appear more favorable. We will consider both deterministic sentiments, where investors add or subtract a constant from all signals, and stochastic sentiments, where sentiment itself is drawn from a distribution.¹¹

An interesting special case is confirmation bias, which is a tendency to misinterpret information that fails to confirm prior beliefs. Suppose that behavioral investors are optimistic, and do not interpret negative information correctly (the case of pessimism is analogous). Paraphrasing the binary model of Rabin and Schrag (1999), assume that they interpret

⁹Equivalently, behavioral traders can only process K signals, where $K = \min\{\max\{k : I(k) \leq \kappa\}, N\}$.

¹⁰This is a reasonable restriction, since inferring information from prices requires the extraction of a sufficient statistic, whose Shannon information is equivalent to that of N signals (Grossman and Stiglitz, 1980), and thus would be beyond the cognitive capacity of traders subject to inattention. Indeed, Kacperczyk et al. (2015) prove formally that learning from prices is not an optimal strategy for a rationally inattentive investor.

¹¹A potential objection to sentiments is that, in a large population of behavioral traders, misperceptions ought to average to zero. However, Hirshleifer (2001) argues that “people share similar heuristics, those that worked well in our evolutionary past. So on the whole, we should be subject to similar biases. Systematic biases (common to most people, and predictable based upon the nature of the decision problem) have been confirmed in a vast literature in experimental psychology” (p. 1540). When modeling sentiments, we are attempting to model these systematic biases.

positive signals $s_i > \theta_0$ correctly, but simply take negative signals $s_i < \theta_0$ to be equal to their prior. Thus the perceived signal is $s_i + x_i = \max\{s_i, \theta_0\}$. This yields the stochastic misperception term $x_i = \max\{0, \theta_0 - s_i\}$, which is a censored Gaussian random variable.

2.2 Equilibrium prices

If stocks trade at price p at date 0, rational investors will demand $x_R = \rho_R \gamma^{-1}(\theta_R - rp)$ units, and behavioral investors will demand $x_B = \rho_B \gamma^{-1}(\theta_B - rp)$ units. The equilibrium price thus solves

$$\begin{aligned} rp &= \frac{\lambda \rho_B \theta_B + (1 - \lambda) \rho_R \theta_R}{\lambda \rho_B + (1 - \lambda) \rho_R} \\ &= q \theta_B + (1 - q) \theta_R. \end{aligned} \tag{7}$$

where $q = \lambda \rho_B / (\lambda \rho_B + (1 - \lambda) \rho_R)$ is a *confidence-weighted* measure of behavioral investors. We now derive general formulae for the average volatility of returns and the average trading volume of trading in equilibrium. We then apply these formulae to our five examples of biases to obtain testable predictions in Section 2.5.

2.3 Volatility

The return on the stock between dates 0 and 1 is $R = \theta/p$. The volatility of returns is $\mathbb{V}[R]$, i.e. the variance of the ratio of the two Gaussian random variables θ and p . There is no closed-form solution for the variance of such a ratio (Hinkley, 1969). For this reason, the theoretical literature on the CARA-Gaussian model focuses on the variance of prices $\mathbb{V}[p]$.

By contrast, the focus of the empirical literature – including our own analysis below – is on the variance of returns $\mathbb{V}[R]$. It is tempting to use $\mathbb{V}[p]$ as a proxy for $\mathbb{V}[R]$, but this is misleading. The variances of prices and returns are not equivalent, and can indeed be negatively correlated.¹²

We address this issue in two ways. First, in this section, we analyze an approximate measure of $\mathbb{V}[R]$ analytically to obtain empirical predictions. Second, in Appendix B, we check numerically that our analytical predictions match the comparative statics of $\mathbb{V}[R]$ when it is computed exactly.

For our approximate measure of variance, assume that the *average* bias in behavioral traders' beliefs is small, so that they approximately satisfy the law of iterated expectations

¹²In a perfectly informed market, for instance, prices always equal the fundamental θ , so that the variance of returns is zero while the variance of prices is large. In an uninformed market, prices always equal the prior θ_0 , so that the variance of prices is zero while the variance of returns is large.

$\mathbb{E}[\theta_B] = \theta_0$. Then the equilibrium price matches the prior on average, $\mathbb{E}[rp] = \theta_0$. Then, a first-order approximation around the unconditional mean gives

$$\frac{\theta_0^2}{r^2} \mathbb{V}[R] \simeq \mathbb{V}[\theta - rp],$$

which we can decompose further by the law of total variance:

$$\mathbb{V}[\theta - rp] = \mathbb{E}[\mathbb{V}[\theta - rp|\mathbf{s}]] + \mathbb{V}[\mathbb{E}[\theta - rp|\mathbf{s}]].$$

Since prices are known conditional on signals, we have $\mathbb{V}(\theta - rp|\mathbf{s}) = \mathbb{V}(\theta|\mathbf{s}) = \rho_R^{-1}$. As for the second term, we have $\mathbb{E}[\theta - rp|\mathbf{s}] = q(\theta_R - \theta_B)$, since rational traders' posterior beliefs match the true state on average. Combining, we obtain the approximate volatility of returns:

$$\frac{\theta_0^2}{r^2} \mathbb{V}[R] \simeq \frac{1}{\rho_0 + N\rho_\epsilon} + q^2 \mathbb{V}[\theta_B - \theta_R] \quad (8)$$

The first term measures the volatility of returns in a fully rational market. This is decreasing in buzz (N): when all traders are rational, more informed markets are less volatile. The second term captures the additional volatility due to behavioral biases. The key statistic is the variance of the *disagreement* between rational and behavioral investors $\mathbb{V}[\theta_B - \theta_R]$, which will also play a key role in determining trading volume. The link between disagreement and buzz depends on the nature of bias which will determine q . We will make this link more precise when we return to our examples below.

2.4 Trading volume

The average trading volume of stocks traded in equilibrium is $T = \|x_B(p^*)\|$. Using the demand function and the equilibrium price in (7), we obtain average trading volume in the market:

$$E[T] = \frac{1}{\gamma} \left[\frac{1}{\lambda\rho_B} + \frac{1}{(1-\lambda)\rho_R} \right]^{-1} \mathbb{E}[\|\theta_B - \theta_R\|]. \quad (9)$$

The effect of buzz on trading volume can be decomposed into two factors. The first is *confidence*: When there are more signals, posterior beliefs are tighter, and the confidence weights q and $1 - q$ are larger. This increases trading volume, since investors bet more aggressively when they are certain of their opinion. The third factor is *average disagreement*. As in [Harris and Raviv \(1993\)](#), “differences of opinion make a horse race”, and markets are more active when traders disagree about the average quality of assets. The influence of disagreement is closely related to the volatility equation (8), but instead of the variance of

Table 1: **Empirical predictions**

	$\frac{\partial \text{Volatility}_{t+1}}{\partial \text{Buzz}_t}$	$\frac{\partial \text{Turnover}_{t+1}}{\partial \text{Buzz}_t}$
Rational market	-	0
1a: Overconfidence, small bias	-	+
1b: Overconfidence, large bias	-	-
2: Stale news	+	+
3: Conservatism	-	+
4: Inattention	-	+
5a: Strong sentiments	+	+
5b: Weak sentiments (incl. confirmation bias)	-	+

returns, it is the expected absolute disagreement which drives trading volume.

As in the volatility equation (8), the effect of buzz on disagreement is ambiguous. If it is positive, then buzz unambiguously increases trading volume. If it is negative, then more buzz can imply less trading volume, so long as the disagreement effect dominates the confidence effect. We now turn to some explicit examples to examine the balance of these effects.

2.5 Empirical predictions

We now revisit our examples of behavioral biases to determine how increases in “buzz”, i.e. in the number of signals N , affects the subsequent volatility of returns and trading volume in each case. As a preview, Table 1 summarizes the predictions in each case:

Rational markets: Consider the case where there are no behavioral traders, $\lambda = 0$. It is immediate from equation (8) that buzz decreases the variance of returns, because it simply renders the market more informative. Moreover, the “no trade theorem” (Milgrom and Stokey, 1982) applies in a rational market, since all investors agree about fundamental values. Hence, more information has no further impact on trading volume, which is always zero.¹³

¹³A common extension to the rational model, which leads to more realistic results on trading volume, is to assume that there are noise traders who generate random demand (Grossman and Stiglitz, 1980; De Long et al., 1991). If noise traders have a downward-sloping demand for assets but do not pay attention to

Example 1: Overconfidence. In the case of overconfidence, disagreement is driven by the excessive weight $\hat{w} > w$ placed on signals by behavioral traders, defined in equation (5). The disagreement between behavioral and rational traders is $\theta_B - \theta_R = N(\hat{w} - w)(\theta_0 - \bar{s})$, where $\bar{s} = N^{-1} \sum_i s_i$ is the average signal. The disagreement is normally distributed with mean zero and variance

$$\mathbb{V}[\theta_B - \theta_R] = N^2(\hat{w} - w)^2 \mathbb{V}[\bar{s}].$$

The absolute value of disagreement $\|\theta_B - \theta_R\|$ has a folded normal distribution, and its expectation is $\mathbb{E}[\|\theta_B - \theta_R\|] = (2\mathbb{V}[\theta_B - \theta_R]/\pi)^{1/2}$. Substituting into (8) and (9) and differentiating, we can sign the effects of buzz on the volatility of returns and average trading volume:

$$\frac{\partial \mathbb{V}[R]}{\partial N} \stackrel{\text{sign}}{=} a\lambda - (1 + a\lambda)^2 \frac{N\rho_\epsilon}{\rho_0} - 1 \quad (10)$$

$$\frac{\partial \mathbb{E}[T]}{\partial N} \stackrel{\text{sign}}{=} (1 - a\lambda) \frac{N\rho_\epsilon}{\rho_0} + 1. \quad (11)$$

When there is a large numbers of signals N , the effect of buzz on market volatility is negative. Intuitively, volatility is reduced by buzz because (i) rational traders become more informed, and (ii) extreme realizations of the average signal, which drive disagreement, become less likely when there are many signals. These effects dominate the fact that behavioral traders become more confident.

The effect of buzz on trading volume for large N is positive when biases are small, $a < 1/\lambda$, and negative when biases are large, $a > 1/\lambda$. Unlike in the volatility equation, the confidence effect can dominate for small biases. This is because the confidence weights of rational *and* behavioral traders both drive up trading volume, while only the confidence of behavioral traders affected variance.

Example 2: Stale news. In the case of stale news, each of M informative signals are repeated $K > 1$ times, and behavioral investors treat the repetitions as new informative signals. Since behavioral traders count each signal K times, this model is formally equivalent to assuming that behavioral traders perceive the precision of signals to be $\hat{\rho}_i = K\rho_\epsilon$. Therefore, the derivation of volatility and trading volume is the same as in the case of overconfidence in Example 1 (defining the overconfidence measure to be $a = K - 1$).

This example is special in that there are two possible interpretations of buzz: We can think of buzz as increasing the number of informative signals M , or as increasing the number

information, then more buzz is likely to reduce trading volume. As the number of signals increase, rational traders become more confident, their demand becomes more elastic, and they hold a larger share of the market. Thus, a market with low buzz will contain a significant amount of buying and selling by noise traders, while a market with high buzz will be well-approximated by the “no trade theorem”.

of repetitions K . If buzz increases the number of informative signals, then its implications are the same as in the case of overconfidence. If buzz increases the number of repetitions, on the other hand, we find that it unambiguously increases both volatility and trading volume:

$$\frac{\partial \mathbb{V}[R]}{\partial K} > 0 \quad (12)$$

$$\frac{\partial \mathbb{E}[T]}{\partial K} > 0. \quad (13)$$

Intuitively, more repetitions increase the disagreement between behavioral and rational traders, but do not increase the precision of the rational traders' beliefs. Hence, the volatility of returns and the average trading volume increase unambiguously.

Example 3: Conservatism. In this case, behavioral traders overstate the weight on their prior, leading to excessively low weights $\hat{w} < w$ on signals, as defined in (6). The steps of analysis are much the same as in the case of overconfidence. We find that buzz unambiguously decreases volatility but increases trading volume:

$$\frac{\partial \mathbb{V}[R]}{\partial N} < 0 \quad (14)$$

$$\frac{\partial \mathbb{E}[T]}{\partial N} > 0. \quad (15)$$

This corresponds to our findings with overconfidence when biases are small. With large biases, trading volume still increases in buzz in the case of conservatism, but falls in the case of overconfidence. This is because disagreement between investors is more resistant to learning when it is driven by conservatism. With overconfidence, the weight placed on the average signal converges to unity as N grows large, both for rational and behavioral investors. With conservatism, the weight placed on the average signal by behavioral investors converges more slowly, since they are wedded to their priors. This effect implies that the average disagreement $\mathbb{E}[\|\theta_B - \theta_R\|]$ decreases less quickly under conservatism, allowing the confidence effect of buzz to dominate.

Example 4: Inattention. With inattention, behavioral traders only process the first $K < N$ signals. Let $\bar{s}_1 = K^{-1} \sum_{i \leq K} s_i$ denote the average signal observed by behavioral traders, and $\bar{s}_2 = (N - K)^{-1} \sum_{i > K} s_i$ the average of the remaining signals. All agents correctly assess the precision of the signals they process, so the weights placed on each signal by behavioral and rational investors are, respectively, $\hat{w} = \rho_\epsilon / (\rho_0 + K \rho_\epsilon)$ and $w = \rho_\epsilon / (\rho_0 + N \rho_\epsilon)$.

Using these weights in (1) and (3) we obtain posterior means:

$$\begin{aligned}\theta_R &= wK\bar{s}_1 + w(N-K)\bar{s}_2 + (1-Nw)\theta_0 \\ \theta_B &= \hat{w}K\bar{s}_1 + (1-\hat{w}K)\theta_0.\end{aligned}$$

The disagreement $\theta_B - \theta_R$ has mean zero, because neither group of investors makes systematic mistakes when processing information, and variance

$$\mathbb{V}[\theta_B - \theta_R] = \rho_\epsilon \frac{N-K}{(\rho_0 + N\rho_\epsilon)(\rho_0 + K\rho_\epsilon)}.$$

As before, the absolute value of disagreement $\|\theta_B - \theta_R\|$ has a folded normal distribution, and its expectation is $\mathbb{E}[\|\theta_B - \theta_R\|] = (2\mathbb{V}[\theta_B - \theta_R]/\pi)^{1/2}$. Substituting into (8) and (9) and differentiating, we find that an increase in the number of signals decreases the variance of returns, but increases trading volume:

$$\frac{\partial \mathbb{V}[R]}{\partial N} < 0 \tag{16}$$

$$\frac{\partial \mathbb{E}[T]}{\partial N} > 0. \tag{17}$$

Example 5: Sentiments and confirmation bias. We have assumed that behavioral traders perceive signals $s_i + x_i$, where x_i captures their sentiment. Beyond this misperception, they process information correctly. Assume for simplicity that all signals have precision ρ_ϵ , so the weight attached to each signal is $w = \rho_\epsilon/(\rho_0 + N\rho_\epsilon)$ for all traders. Suppose that for all signals i , the error is normally distributed $x_i \sim \mathcal{N}(\mu_x, \sigma_x^2)$, and that the correlation between two errors is $\text{Corr}(x_i, x_j) = \rho_x$. The disagreement between rational and behavioral traders is $\theta_B - \theta_R = Nw\bar{x}$, where \bar{x} is the average error. Thus the disagreement has mean $Nw\mu_x \equiv \mu_\Delta$ and variance

$$\mathbb{V}[\theta_B - \theta_R] = N^2w^2\sigma_x^2 \left[\frac{1}{N} + \left(1 - \frac{1}{N}\right)\rho_x \right] \equiv \sigma_\Delta^2.$$

The absolute value of disagreement has a folded normal distribution, and its expectation is given by

$$\mathbb{E}[\|\theta_B - \theta_R\|] = \sqrt{\frac{2\sigma_\Delta^2}{\pi}} \exp\left(-\frac{\mu_\Delta^2}{2\sigma_\Delta^2}\right) + \mu_\Delta \left[1 - 2\Phi\left(-\frac{\mu_\Delta}{\sigma_\Delta}\right)\right].$$

Substituting into (8) and (9) and differentiating, we can sign the effect of buzz on volatility

and trading volume when there is a large number of signals:

$$\lim_{N \rightarrow \infty} \frac{\partial \mathbb{V}[R]}{\partial N} \stackrel{\text{sign}}{=} \lambda^2 \sigma_x^2 [2\rho_x \rho_0 - (1 - \rho_x)\rho_\epsilon] - 1. \quad (18)$$

$$\lim_{N \rightarrow \infty} \frac{\partial \mathbb{E}[T]}{\partial N} > 0. \quad (19)$$

When there are many signals, buzz increases trading volume. Buzz increases volatility when the variance of errors σ_x^2 and the correlation between errors ρ_x are both sufficiently large. We will call this the case of “strong sentiments”: In this parametric region, sentiments have a robust impact on behavioral investors’ trading activity, because (i) they have a sizable impact on the volatility of their beliefs, and (ii) the errors made when interpreting different signals do not cancel out, because they are correlated with one another. For example, if the misperception induced by sentiments is the same for each signal, we have $\rho_x = 1$, and the strong sentiments case is $\lambda^2 \sigma_x^2 > \frac{1}{2\rho_0}$, i.e. when the variance of sentiments, weighted by the population mass of behavioral traders, is at least half the (prior) variance of the true asset value.

In the special case of confirmation bias (with optimistic investors), behavioral errors occur for signals below the prior $s_i < \theta_0$, which are misperceived to be equal to the prior, yielding $x_i = \max\{0, \theta_0 - s_i\}$. Thus, errors are censored normal variables. In Appendix A we derive the variance and correlation of errors:

$$\begin{aligned} \sigma_x^2 = \mathbb{V}[x_i] &= \left(\frac{1}{\rho_0} + \frac{1}{\rho_\epsilon} \right) \left(\frac{1}{2} - \frac{1}{2\pi} \right) \\ \rho_x = \text{Corr}[x_i, x_j] &= \left(\frac{1}{2} - \frac{1}{2\pi} \right)^{-1} \left[\zeta(r)r - \frac{1}{2\pi} (1 - \sqrt{1 - r^2}) \right], \end{aligned}$$

where $r = \rho_\epsilon / (\rho_0 + \rho_\epsilon)$ is the correlation between two signals s_i and s_j , and

$$\zeta(r) = \frac{1}{4} + \frac{1}{2\pi} \text{ArcSin}(r)$$

denotes the probability that two signals both lie below the prior. Substituting into (18), we find that buzz increases volatility for a large number of signals if and only if

$$\frac{2 - r}{r(1 - r)} \left[\zeta(r)r - \frac{1}{2\pi} (1 - \sqrt{1 - r^2}) \right] - \frac{1}{1 - r} \left(\frac{1}{2} - \frac{1}{2\pi} \right) > 1.$$

This inequality is difficult to solve analytically, but numerical simulations suggest that the maximum of the left-hand side, over the possible range of correlations $r \in [0, 1]$, is less than 0.25. Thus, it appears that buzz decreases volatility when there is confirmation bias and a

sufficiently large number of signals.

Considering trading volume, we cannot use the expression in (19), since the disagreement between behavioral and rational traders $\theta_B - \theta_R = Nw\bar{x}$ is not normally distributed due to censoring of the average error $\bar{x} = N^{-1} \sum_i \max\{0, \theta_0 - s_i\}$. However, since behavioral investors are optimistic in this example, they only over-interpret signals, and so the average error is positive by definition. Hence, using the properties of censored normal variables, and substituting into (9), it follows that trading volume with confirmation bias is a linear increasing function of N :

$$\mathbb{E}[T] = N \frac{(1-\lambda)\lambda}{\gamma} \sqrt{\frac{1}{2\pi} \left(\frac{1}{\rho_0} + \frac{1}{\rho_\epsilon} \right)}.$$

To sum up, confirmation bias implies that

$$\lim_{N \rightarrow \infty} \frac{\partial \mathbb{V}[R]}{\partial N} < 0. \quad (20)$$

$$\frac{\partial \mathbb{E}[T]}{\partial N} > 0. \quad (21)$$

This corresponds to the general case of “weak sentiments”. With confirmation bias, the variance and correlation of perceptive errors are not large enough to generate a positive effect of buzz on volatility. Nonetheless, buzz does increase the average disagreement and therefore drives up trading volume.

Equations (10) through (21) constitute our central empirical predictions, which are also collated in Table (1). The remainder of this paper tests these predictions using our data on social and news media buzz.

3 Data

This section describes the data we use to measure stock coverage in social and news media, as well as financial data we use to measure stock prices, volatility, trading volume and characteristics.

3.1 Measuring ‘buzz’ in social and news media

Our main results concern the level of coverage in social and news media – or “buzz” – enjoyed by different stocks. We use the Thompson Reuters MarketPsych Index (TRMI) database, which extracts measures of buzz and sentiment from English-language news and social media content using a machine learning lexical analysis algorithm. The social and news media

sources covered by the algorithm have evolved over time. For consistency, we focus on the period between January 2009 and December 2014.

During this time, the main sources of news media content are (i) Reuters News, (ii) a host of mainstream news sources collected by MarketPsych Data, and (iii) online content collected by Moreover Technologies from about 50,000 internet news sites. The online news content includes many finance-specific sites such as Forbes and SeekingAlpha.

The main source of social media content is a social media feed constructed by Moreover Technologies. This captures the top 30% of social media content, as measured by popularity using incoming links, collected from around 4 million social media sources such as chat rooms (including stock-market specific chats), Facebook posts, blogs and tweets.

From these sources, the TRMI algorithm extracts high-frequency measures of buzz surrounding each of about 3000 US stocks. The *total buzz* of a stock is on a given day counts the number of words referring to the stock in the above sources. This number is obtained by first identifying articles and social media posts referring to events (for instance, litigation, mergers or layoffs) related to a stock, and then counting the length of those articles. Total buzz is divided by the total buzz of all stocks mentioned on the day to obtain *relative buzz*.

This calculation is done separately for social and news media content, yielding our key measures of coverage: Social media relative buzz (BuzzS) and news media relative buzz (BuzzN), both of which are continuous variables between zero and one. We convert these daily measures of buzz to monthly averages for our main analysis.

We focus on stocks which are traded on NYSE, AMEX and NASDAQ. As is standard in this kind of empirical analysis, we exclude regulated utilities (SIC codes 4910-4949), depository institutions (SIC 6000-6099) and holding and investment companies (SIC 6700-6799). This yields a panel of 2613 stocks, observed between January 2009 and December 2014. The panel is unbalanced due to the entry and exit of stocks, and includes 162,268 stock-month observations. In Section 4.3, we show that our main results are preserved when we restrict attention to a balanced panel.

3.2 Financial data

We merge our measures of buzz with monthly financial data from the Center for Research in Securities Prices (CRSP) and the Compustat database. The main variables of interest are the trading volume and the realized idiosyncratic volatility of each stock. Our measure of trading volume (Turn, for “turnover”) is taken directly from the CRSP data.

Our parametric measure of realized volatility (iVolp) is constructed in two steps. First, for every month m in the model, we estimate a three-factor model of daily returns on each

stock, by running the regression

$$(R_{it} - R_{ft}) = +\beta_1^{(m)}(Rm_t - R_{ft}) + \beta_2^{(m)}SMB_t + \beta_3^{(m)}HML_t + \epsilon_t^{(m)},$$

where R_{it} is the return to stock i on day t , R_{ft} is the one-month treasury bill rate; Rm_t is the return to the value-weighted market portfolio; SMB_t is the average return on the three small-cap portfolios minus the average return on the three big-cap portfolios; and HML_t is the average return on the two value stock portfolios minus the average return on the two growth stock portfolios. Second, we define the idiosyncratic volatility of stock i in month m as the sum of squared errors from this monthly regression. As a robustness check, in Section 4.3, we also use a non-parametric measure of idiosyncratic volatility (iVoln), which is obtained by taking the variance of returns of each stock at the monthly frequency.

Our analysis also includes a set of financial variables which have predictive power for volatility and trading volume. We obtain measures of firm size (Size), monthly stock price returns (Ret), and the standard deviation of the last 60 monthly returns (TotalSD) from CRSP. Using the Compustat data, we calculate each firm’s leverage (Leverage), the book-to-market ratio (BM) and its degree of ‘focus’ as measured by the Herfindahl-Hirschman index of segment revenue (HHI). Finally, we include the share of institutional ownership (InstOwn) from the Thomson Reuters Stock Ownership Summary.¹⁴

3.3 Summary statistics

Table 2 shows sample means and standard deviations for buzz, volatility and trading volume, both for the total stocks and disaggregated by industry. Buzz is measured in percentage points. We winsorize all financial data and the buzz measures at 1% to ensure that our estimates are robust to the effect of outliers.

The sample average of relative buzz is equal to about 0.02%. Values of buzz range from zero for stocks which are not covered in a given month to 0.5% for the most talked about stocks.¹⁵ The standard deviation of buzz slightly higher for news media than for social media. There is evidence of heterogeneity across industries: Buzz is relatively high in the trade, services, manufacturing and finance industries, and relatively low in agriculture and mining/construction.

Table 3 reports sample averages and standard deviations for our financial control vari-

¹⁴The literature on stock market volatility demonstrates the predictive power of size (Cheung and Ng, 1992), returns (Duffee, 1995), institutional ownership (Dennis and Strickland, 2002) and trading volume (Schwert, 1989). trading volume is commonly associated with the absolute value of returns (Karpoff, 1987; Schwert, 1989), institutional ownership (Tkac, 1999) and size (Tkac, 1999; Lo and Wang, 2000).

¹⁵Before we winsorize the top 1%, we observe stocks which have relative buzz of about 10%.

Table 2: **Summary statistics for buzz, volatility and turnover**

Industries		BuzzN	BuzzS	iVolp	Turn
Agriculture	Mean	0.0118	0.0074	0.0108	1.4236
	SD	0.0197	0.0102	0.0179	1.8153
	Max	0.2315	0.0687	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	520	520	520	520
Mining&Constr	Mean	0.0174	0.0264	0.0131	3.0775
	SD	0.0314	0.0688	0.0198	2.5168
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	11106	11106	11106	11106
Manufacturing	Mean	0.0288	0.0314	0.0111	1.9947
	SD	0.0657	0.0897	0.0197	1.8134
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	71569	71569	71569	71569
Utilities	Mean	0.0330	0.0240	0.0107	1.9643
	SD	0.0793	0.0667	0.0202	1.7116
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	9705	9705	9705	9705
Trade	Mean	0.0323	0.0225	0.0093	2.1809
	SD	0.0657	0.0647	0.0164	1.7985
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	17709	17709	17709	17709
Finance	Mean	0.0311	0.0229	0.0073	1.5866
	SD	0.0685	0.0781	0.0156	1.5239
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	13376	13376	13376	13376
Services	Mean	0.0327	0.0304	0.0111	1.9164
	SD	0.0803	0.0965	0.0190	1.7618
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	30908	30908	30908	30908
Other	Mean	0.0239	0.0270	0.0144	1.8484
	SD	0.0664	0.0834	0.0205	1.9401
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	7375	7375	7375	7375
Total	Mean	0.0293	0.0285	0.0109	2.0303
	SD	0.0682	0.0849	0.0191	1.8644
	Max	0.5323	0.6500	0.1310	11.0055
	Min	0.0000	0.0000	0.0003	0.1062
	N	162268	162268	162268	162268

ables, and Table 4 shows contemporaneous correlations, with standard errors in parentheses. Social and news media buzz are heavily correlated, as expected. Buzz correlates with firm size, and this correlation is stronger for news than for social media. There is also a strong correlation between buzz and trading volume, especially for social media. The contemporaneous correlation between volatility and buzz is positive for social media, and negative (but smaller in magnitude) for news media.

Table 3: **Summary statistics for control variables**

	Size	InstOwn	Ret	HHI	Leverage	TotalSD
Mean	13.8058	0.6972	0.0196	0.7389	0.0258	0.1430
SD	1.7366	0.2438	0.1274	0.2923	0.0546	0.0632
Max	18.3953	1.0000	0.4822	1.0000	0.3417	0.3791
Min	10.1016	0.0597	-0.3136	0.0000	0.0000	0.0489
N	159358	162268	161526	141532	151185	162220

Table 4: **Contemporaneous correlations**

Variables	BuzzN	BuzzS	iVolp	Turn	Size	InstOwn	Ret	HHI	Leverage	TotalSD
BuzzN	1.000									
BuzzS	0.554 (0.000)	1.000								
iVolp	-0.032 (0.000)	0.100 (0.000)	1.000							
Turn	0.133 (0.000)	0.307 (0.000)	0.258 (0.000)	1.000						
Size	0.409 (0.000)	0.230 (0.000)	-0.386 (0.000)	0.176 (0.000)	1.000					
InstOwn	0.012 (0.000)	-0.085 (0.000)	-0.174 (0.000)	0.235 (0.000)	0.339 (0.000)	1.000				
Ret	-0.001 (0.691)	-0.010 (0.000)	0.120 (0.000)	0.023 (0.000)	-0.042 (0.000)	-0.010 (0.000)	1.000			
HHI	-0.076 (0.000)	-0.008 (0.005)	0.076 (0.000)	0.049 (0.000)	-0.173 (0.000)	-0.052 (0.000)	0.003 (0.242)	1.000		
Leverage	0.059 (0.000)	0.051 (0.000)	0.065 (0.000)	-0.008 (0.002)	-0.027 (0.000)	-0.044 (0.000)	0.007 (0.006)	-0.026 (0.000)	1.000	
TotalSD	-0.132 (0.000)	0.077 (0.000)	0.331 (0.000)	0.185 (0.000)	-0.470 (0.000)	-0.166 (0.000)	0.020 (0.000)	0.109 (0.000)	0.077 (0.000)	1.000

4 Empirical results

In Section 2, we developed a general model of asset pricing and information processing. Different specifications of behavioral biases resulted in different implications for the relationship between buzz and subsequent idiosyncratic volatility and trading volume, which are summarized in Table 1 on page 10. We now test these implications by examining whether buzz predicts volatility and trading volume.

4.1 Idiosyncratic volatility

Let i index stocks and t index months. We run a panel regression of stock i 's next month's volatility $iVolp_{i,t+1}$ on this month's buzz in social and news media, $BuzzS_{i,t}$, stock-level control variables $X_{i,t}$, and this month's volatility $iVolp_{i,t}$. We control for stock and time fixed effects α_i and μ_t :

$$iVolp_{i,t+1} = \alpha_i + \mu_t + \beta_S \times BuzzS_{i,t} + \beta_N \times BuzzN_{i,t} + \gamma \cdot X_{i,t} + \delta \times iVolp_{i,t} + \epsilon_{i,t+1}. \quad (22)$$

Table 5 reports the results from estimating (22).¹⁶

We consistently find that high news media buzz predicts lower future volatility and this is statistically significant at the 1% level. Social media buzz predicts higher future volatility is equally significant once we control for stock characteristics in column (3). The sign of the coefficient on social media buzz changes once we add the controls, suggesting that the estimates in columns (1) and (2) were biased upwards by omitted variables. Including time fixed effects in column (4) makes little difference, suggesting that the estimated effects of buzz are not driven by unobserved time trends.

To interpret the economic significance of buzz, suppose that a stock goes from having no buzz in social media to being one of the most talked-about stocks with a relative buzz of 0.5%. Then, according to the most general specification (column (4)), our measure of the stock's subsequent idiosyncratic volatility rises by about 0.0055 on average. This increase corresponds to about half the average volatility, or just under a third of a standard deviation. For an equivalent change in news media buzz, the stock's subsequent volatility falls by 0.008, about a three quarters of the average volatility, or 40% of a standard deviation.

¹⁶Standard errors are clustered at the stock level, and the corresponding t-statistics are in parentheses. Stock fixed effects are included throughout to ensure that our estimates are not the result of unobserved stock-level characteristics.

Table 5: **Volatility regressions**

	(1)	(2)	(3)	(4)
	iVolp(+1)	iVolp(+1)	iVolp(+1)	iVolp(+1)
iVolp	0.302*** (34.70)	0.225*** (27.52)	0.213*** (24.20)	0.185*** (21.21)
BuzzN	-0.0199*** (-7.86)	-0.0187*** (-8.19)	-0.0193*** (-7.44)	-0.0167*** (-6.85)
BuzzS	-0.00524** (-2.23)	0.000559 (0.24)	0.00991*** (4.15)	0.0113*** (4.85)
Size			-0.00791*** (-34.30)	-0.00618*** (-21.70)
InstOwn			0.00462*** (4.31)	0.00298*** (2.70)
HHI			0.000211 (0.31)	-0.0000732 (-0.11)
Leverage			0.00282 (1.24)	0.00354 (1.52)
Turn			-0.000235*** (-2.88)	-0.000317*** (-3.59)
Ret			-0.0108*** (-19.29)	-0.00952*** (-14.77)
TotalSD			-0.0769*** (-13.43)	-0.0627*** (-9.98)
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
N	159549	159549	131878	131878
R^2	0.092	0.159	0.142	0.178

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.2 Trading volume

For trading volume, we run the analogous panel regression:

$$Turn_{i,t+1} = \alpha_i + \mu_t + \beta_S \times BuzzS_{i,t} + \beta_N \times BuzzN_{i,t} + \gamma \cdot X_{i,t} + \delta \times Turn_{i,t} + \epsilon_{i,t+1}. \quad (23)$$

Table 6 reports our estimates.¹⁷

We consistently find that high social media buzz predicts high trading volume, and that high news media buzz predicts low trading volume. This effect is statistically significant at the 1% level, and does not change much when we include control variables in column (3), or time fixed effects in columns (2) and (4).

To interpret the effects, suppose again that relative social media buzz around a stock rises from zero to the level of the most talked-about stocks at 0.5%. Then according to the most general specification in column (4), trading volume increases by about 1/2, which corresponds to 24% of average trading volume, or about a quarter of a standard deviation. For an equivalent rise in news media buzz, trading volume falls by about 0.88, which is 43% of average trading volume, or 47% of a standard deviation.

4.3 Robustness

To examine the robustness of our results, we run the volatility and trading volume regressions (22) and (23) on a balanced panel of stocks, and using the non-parametric measure of volatility $iVoln_{i,t}$.

Tables 7 and 8 show the results on a balanced panel. The estimated effects of buzz on volatility and trading volume are quantitatively very similar to those obtained in the baseline specification. Moreover, the pattern of statistical significance is unchanged. After controlling for stock-level characteristics, the effects of social and news media buzz on volatility are significant at the 1% level. The corresponding effects on trading volume are also significant at the 1% level throughout. All estimates are robust to the inclusion of stock and time fixed effects.

Table 9 shows the results of running the regression (22) using the non-parametric measure of volatility. Again, the estimated effects of buzz are quantitatively similar to the baseline model, and significant at the 1% level.

Tables 7 to 9 thus suggest that our estimated effects of buzz on volatility and trading volume are not driven by our choice of volatility measure, or by the entry and exit of stocks to and from the sample.

¹⁷Standard errors are clustered at the stock level, and the corresponding t-statistics are in parentheses. Again, we control for stock fixed effects throughout.

Table 6: **Turnover regressions**

	(1)	(2)	(3)	(4)
	Turn(+1)	Turn(+1)	Turn(+1)	Turn(+1)
Turn	0.584*** (89.66)	0.590*** (88.03)	0.589*** (75.51)	0.595*** (74.12)
BuzzN	-1.841*** (-7.21)	-1.974*** (-7.71)	-1.670*** (-6.03)	-1.765*** (-6.38)
BuzzS	0.865*** (5.23)	0.852*** (5.21)	1.014*** (5.21)	0.961*** (4.96)
Size			-0.0259 (-1.64)	0.0628*** (3.22)
InstOwn			0.431*** (6.34)	0.304*** (4.44)
HHI			0.0384 (0.82)	0.00204 (0.04)
Leverage			0.123 (0.98)	0.0559 (0.45)
iVolp			-5.655*** (-14.37)	-6.107*** (-15.46)
AbsRet			0.498*** (9.78)	0.395*** (7.78)
TotalSD			-0.953*** (-3.18)	-0.435 (-1.30)
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
N	159613	159613	131891	131891
R^2	0.346	0.384	0.340	0.380

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 7: **Volatility regressions with balanced panel**

	(1)	(2)	(3)	(4)
	iVolp(+1)	iVolp(+1)	iVolp(+1)	iVolp(+1)
iVolp	0.338*** (33.73)	0.256*** (26.53)	0.241*** (24.19)	0.212*** (21.36)
BuzzN	-0.0180*** (-7.08)	-0.0175*** (-7.55)	-0.0182*** (-7.46)	-0.0158*** (-6.75)
BuzzS	-0.00445** (-2.01)	0.00125 (0.58)	0.00898*** (3.72)	0.0104*** (4.34)
Controls	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
N	131208	131208	112515	112515
R^2	0.117	0.184	0.163	0.197

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 8: **Turnover regressions with balanced panel**

	(1)	(2)	(3)	(4)
	Turn(+1)	Turn(+1)	Turn(+1)	Turn(+1)
Turn	0.609*** (86.29)	0.614*** (83.46)	0.613*** (75.35)	0.619*** (73.07)
BuzzN	-1.791*** (-7.67)	-1.956*** (-8.45)	-1.655*** (-6.38)	-1.755*** (-6.71)
BuzzS	0.720*** (4.25)	0.717*** (4.28)	0.920*** (4.71)	0.852*** (4.36)
Controls	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
N	131208	131208	112515	112515
R^2	0.373	0.416	0.371	0.415

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Table 9: **Volatility Regressions with non-parametric measure of volatility**

	(1)	(2)	(3)	(4)
	iVoln(+1)	iVoln(+1)	iVoln(+1)	iVoln(+1)
iVoln	0.407*** (52.81)	0.282*** (35.83)	0.321*** (41.23)	0.247*** (29.70)
BuzzN	-0.0327*** (-8.07)	-0.0281*** (-8.37)	-0.0337*** (-7.81)	-0.0268*** (-7.28)
BuzzS	-0.0119*** (-3.64)	0.00179 (0.55)	0.0103*** (2.90)	0.0154*** (4.52)
Controls	Yes	Yes	Yes	Yes
Stock FE	Yes	Yes	Yes	Yes
Month FE	No	Yes	No	Yes
N	159655	159655	131922	131922
R^2	0.171	0.312	0.213	0.326

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

4.4 Discussion

We can interpret our empirical results using the predictions of our model in Section 2. We have found a persistent pattern: Buzz in social media predicts high idiosyncratic volatility and high trading volume, while buzz in news media predicts low volatility and low trading volume. Since both effect go in opposite directions, we can reject the hypothesis that the processing of information from social and news media is described by the same underlying model. We can further interpret our results by comparing them to the empirical predictions in 1 on page 10.

Regarding social media, we find that the empirical evidence is inconsistent with a rational market model, but consistent with two models of boundedly rational information processing: strong random sentiments, and stale news. A model of strong random sentiments can explain our results by assuming that investors' interpretation of social media content are subject to significant and volatile misperceptions. A model of stale news can explain the results based on the premise that social media content does not reflect new information. Thus, our results lend support to Tetlock's (2011) hypothesis that stale news are misinterpreted, and further suggest that these effects are most pronounced when signals originate from social media.

Regarding news media, our findings are inconsistent with the basic rational model. The effect of buzz on volatility is negative, which is consistent with the rational theory. However,

the effect of buzz on trading volume is also negative, which cannot be explained by a rational market where trading volume is always zero due to a “no trade theorem”. Within the class of models we consider, the results on news media are consistent with a market where some investors are overconfident, and where this bias is large. When more signals arrive, investors make inferences from a larger sample of information. This reduces the variance of beliefs and the likelihood of extreme disagreements, and consequently decreases volatility and trading volume.

We now turn to the question whether our results could be described by a fully rational model which goes beyond our baseline setup. The first possibility is to introduce noise traders, whose demand is random and who do not pay attention to information. It seems plausible that noise trading would lead to higher trading volume in markets where informed traders have a lot of information: When more information arrives, informed traders become more confident in their beliefs, and trade more aggressively to exploit the mispricing generated by noise traders. Thus, a rational model with noise traders is likely to imply that more information (more buzz) would lead to reduced volatility and increased trading volume, since it renders market prices more accurate but encourages trading via the confidence effect. Such a model is therefore unlikely to match our findings on either social or news media.

A second extension is to recognize that buzz – especially in social media – is endogenously generated. Casual empiricism and introspection suggest that people are most likely to post on social media if they disagree with each other. Moreover, large investors might strategically post their opinions on social media after placing an order, hoping that doing so will move markets in their favor. Therefore, an exogenous shock which increases disagreement among traders would be able to explain both a rise in buzz, and a rise in subsequent volatility and trading volume.

The possibility of endogenous buzz is an important caveat to the class of models we consider, which we plan to address in more detail in future work. However, it is unlikely that all our results are driven by the endogeneity of buzz. With endogenous buzz, the sequence of events is different from what we observed. In that case, a disagreement shock would occur first, and then affect market activity and buzz contemporaneously. By contrast, our regressions show that even *controlling* for one month’s volatility and trading volume, one can predict the following month’s activity using buzz in social media.

5 Conclusions

This paper has used a new dataset on news and social media coverage, or “buzz”, surrounding individual stocks. We have established two robust facts about the relationship between such

coverage and market activity. First, intense social media coverage predicts high volatility of returns and high trading volume over the next month. Second, intense news media coverage predicts low volatility and low trading volume. These effects are statistically and economically significant. The difference between the impact of social and news media is striking, and new to the literature on asset pricing and information.

We have related these results to the predictions of our asset pricing model, which allows for both rational investors and a variety of behavioral biases. The evidence on news media is consistent with a model in which some investors are overconfident when interpreting news. The evidence on social media is inconsistent with rational markets. In the class of models we consider, it is consistent with the hypothesis that investors react to “stale news” (the repetition of old facts on social media). It is also consistent with exogenous but random sentiments among investors. Arguably, the stale news hypothesis has greater intuitive appeal because (i) it can explain both our results and the findings of Tetlock’s (2011), and (ii) casual empiricism suggests that much of social media activity is indeed based on sharing and re-tweeting existing content. Thus, our findings suggest that further testing the implications of stale news using social media data is a promising avenue for research.

References

- Alpert, Marc and Howard Raiffa**, “A Progress Report on the Training of Probability Assessors,” in Daniel Kahneman, Paul Slovic, and Amos Tversky, eds., *Judgment Under Uncertainty: Heuristics and Biases*, Cambridge University Press, 1982, pp. 294–305.
- Antweiler, Werner and Murray Z Frank**, “Is all that talk just noise? The information content of internet stock message boards,” *The Journal of Finance*, 2004, 59 (3), 1259–1294.
- Baker, Malcolm and Jeffrey Wurgler**, “Investor sentiment in the stock market,” 2007.
- Barber, Brad M and Terrance Odean**, “All that glitters: The effect of attention and news on the buying behavior of individual and institutional investors,” *Review of Financial Studies*, 2008, 21 (2), 785–818.
- Barberis, Nicholas and Richard Thaler**, “A survey of behavioral finance,” *Handbook of the Economics of Finance*, 2003, 1, 1053–1128.
- , **Andrei Shleifer, and Robert Vishny**, “A model of investor sentiment,” *Journal of financial economics*, 1998, 49 (3), 307–343.
- Bushee, Brian J, John E Core, Wayne Guay, and Sophia JW Hamm**, “The role of the business press as an information intermediary,” *Journal of Accounting Research*, 2010, 48 (1), 1–19.
- Chen, Hailiang, Prabuddha De, Yu Jeffrey Hu, and Byoung-Hyoun Hwang**, “Wisdom of crowds: The value of stock opinions transmitted through social media,” *Review of Financial Studies*, 2014, 27 (5), 1367–1403.
- Cheung, Yin-Wong and Lilian K. Ng**, “Stock Price Dynamics and Firm Size: An Empirical investigation,” *The Journal of Finance*, 1992, 47 (5), 1985–1997.
- Coval, Joshua D and Tyler Shumway**, “Is sound just noise?,” *The Journal of Finance*, 2001, 56 (5), 1887–1910.
- Cutler, David M, James M Poterba, and Lawrence H Summers**, “What moves stock prices?,” *The Journal of Portfolio Management*, 1989, 15 (3), 4–12.
- Da, Zhi, Joseph Engelberg, and Pengjie Gao**, “The sum of all FEARS: Investor sentiment and asset prices,” *Review of Financial Studies*, 2015, 28 (1), 1–32.

- Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam**, “Investor psychology and security market under-and overreactions,” *the Journal of Finance*, 1998, *53* (6), 1839–1885.
- De Bondt, Werner FM and Richard Thaler**, “Does the stock market overreact?,” *The Journal of finance*, 1985, *40* (3), 793–805.
- De Long, J Bradford, Andrei Shleifer, Lawrence H Summers, and Robert J Waldmann**, “The Survival of Noise Traders in Financial Markets,” *Journal of Business*, 1991, pp. 1–19.
- DeMarzo, Peter M., Dimitri Vayanos, and Jeffrey Zwiebel**, “Persuasion Bias, Social Influence, and Unidimensional Opinions,” *The Quarterly Journal of Economics*, 2003, *118* (3), 909–968.
- Dennis, Patrick J. and Deon Strickland**, “Who Blinks in Volatile Markets, Individuals or Institutions?,” *The Journal of Finance*, 2002, *57* (5), 1923–1949.
- Duffee, Gregory R.**, “Stock returns and volatility a firm-level analysis,” *Journal of Financial Economics*, 1995, *37* (3), 399–420.
- Edwards, Ware**, “Conservatism in human information processing,” in Benjamin Kleinmütz, ed., *Formal Representation of Human Judgment*, John Wiley and Sons, 1968, pp. 17–52.
- Fang, Lily and Joel Peress**, “Media coverage and the cross-section of stock returns,” *The Journal of Finance*, 2009, *64* (5), 2023–2052.
- Fischhoff, Baruch, Paul Slovic, and Sarah Lichtenstein**, “Knowing with certainty: The appropriateness of extreme confidence.,” *Journal of Experimental Psychology: Human perception and performance*, 1977, *3* (4), 552.
- Gervais, Simon and Terrance Odean**, “Learning to be overconfident,” *Review of Financial studies*, 2001, *14* (1), 1–27.
- Greene, William H.**, *Econometric Analysis*, Pearson, 2012.
- Griffin, Dale and Amos Tversky**, “The weighing of evidence and the determinants of confidence,” *Cognitive psychology*, 1992, *24* (3), 411–435.
- Grossman, Sanford J. and Joseph E. Stiglitz**, “On the Impossibility of Informationally Efficient Markets,” *American Economic Review*, 1980, *70* (3), 393–408.

- Harris, Milton and Artur Raviv**, “Differences of opinion make a horse race,” *Review of Financial studies*, 1993, 6 (3), 473–506.
- Hinkley, D. V.**, “On the ratio of two correlated normal random variables,” *Biometrika*, 1969, 56 (3), 635–639.
- Hirshleifer, David**, “Investor psychology and asset pricing,” *The Journal of Finance*, 2001, 56 (4), 1533–1597.
- Kacperczyk, Marcin, Jaromir Nosal, and Luminita Stevens**, “Investor Sophistication and Capital Income Inequality,” 2015. Mimeo.
- Kahneman, Daniel and Amos Tversky**, “On the psychology of prediction,” *Psychological review*, 1973, 80 (4), 237.
- Karpoff, Jonathan M.**, “The relation between price changes and trading volume: A survey,” *Journal of Financial and quantitative Analysis*, 1987, 22 (01), 109–126.
- Kyle, Albert S and F Albert Wang**, “Speculation duopoly with agreement to disagree: Can overconfidence survive the market test?,” *The Journal of Finance*, 1997, 52 (5), 2073–2090.
- Lo, Andrew W and Jiang Wang**, “Trading volume: definitions, data analysis, and implications of portfolio theory,” *Review of Financial Studies*, 2000, 13 (2), 257–300.
- Milgrom, Paul and Nancy Stokey**, “Information, trade and common knowledge,” *Journal of Economic Theory*, 1982, 26 (1), 17–27.
- Muthen, Bengt**, “Moments of the censored and truncated bivariate normal distribution,” *British Journal of Mathematical and Statistical Psychology*, 1990, 43 (1), 131–143.
- Odean, Terrance**, “Volume, volatility, price, and profit when all traders are above average,” *The Journal of Finance*, 1998, 53 (6), 1887–1934.
- Peng, Lin and Wei Xiong**, “Investor attention, overconfidence and category learning,” *Journal of Financial Economics*, 2006, 80 (3), 563–602.
- Pouget, Sebastien, Julien Sauvagnat, and Stephane Villeneuve**, “A Mind is a Terrible Thing to Change: Confirmatory Bias in Financial Markets,” 2014. Mimeo.
- Rabin, Matthew and Joel L Schrag**, “First impressions matter: A model of confirmatory bias,” *Quarterly Journal of Economics*, 1999, 114 (1), 37–82.

- Scheinkman, José A. and Wei Xiong**, “Overconfidence and Speculative Bubbles,” *Journal of Political Economy*, 2003, *111* (6), 1183–1220.
- Schwert, G. William**, “Why Does Stock Market Volatility Change Over Time?,” *The Journal of Finance*, 1989, *44* (5), 1115–1153.
- Sims, Christopher A.**, “Implications of rational inattention,” *Journal of Monetary Economics*, 2003, *50* (3), 665–690.
- Tetlock, Paul C.**, “Giving content to investor sentiment: The role of media in the stock market,” *The Journal of Finance*, 2007, *62* (3), 1139–1168.
- , “All the news that’s fit to reprint: Do investors react to stale information?,” *Review of Financial Studies*, 2011, *24* (5), 1481–1512.
- Tkac, Paula A.**, “A trading volume benchmark: Theory and evidence,” *Journal of Financial and Quantitative Analysis*, 1999, *34* (01), 89–114.

A Derivations with confirmation bias

Letting $x_i = \max\{\theta_0 - s_i, 0\}$, we compute $Var(x_i)$ and $Cov(x_i, x_j)$. Using the fact that x_i is a censored version of the normal variable $y_i \equiv \theta_0 - s_i \sim N(0, \rho_0^{-1} + \rho_\epsilon^{-1})$, we obtain¹⁸

$$Var(x_i) = \left(\frac{1}{\rho_0} + \frac{1}{\rho_\epsilon} \right) \left(\frac{1}{2} - \frac{1}{2\pi} \right). \quad (24)$$

Also note that $E[x_i] = \sqrt{\left(\frac{1}{\rho_0} + \frac{1}{\rho_\epsilon} \right) \frac{1}{2\pi}}$. For the covariance, we also require the product expectation

$$\begin{aligned} E[x_i x_j] &= Pr[x_i > 0, x_j > 0] E[x_i x_j | x_i > 0, x_j > 0] \\ &= Pr[x_i > 0, x_j > 0] E[y_i y_j | y_i > 0, y_j > 0], \end{aligned}$$

where the first line uses the fact that $x_i x_j = 0$ whenever one of x_i and x_j is negative. Let $\zeta = Pr[x_i > 0, x_j > 0]$. Then $E[x_i x_j]$ is equal to ζ times the expected product of two normal variables, y_i and y_j , truncated over the region where both are positive. Define the variance-covariance matrix

$$Cov(y_i, y_j) = \sigma^2 \begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix} \equiv R, \quad (25)$$

where $r \equiv Corr(s_i, s_j) = \frac{\rho_\epsilon}{\rho_0 + \rho_\epsilon}$. The joint moment-generating function of $y = (y_i, y_j)$ given that $\{y_i > 0, y_j > 0\}$, denoted $M(t)$ for a real vector $t = (t_i, t_j)$, is defined by

$$\begin{aligned} \zeta M(t) &= e^{\frac{1}{2} t' R t} \times \int_{y > 0} \frac{1}{2\pi |R|} e^{-\frac{1}{2} (y - R t)' R^{-1} (y - R t)} dy \\ &= e^{\frac{1}{2} t' R t} \times \int_{z > -R t} \frac{1}{2\pi |R|} e^{-\frac{1}{2} z' R^{-1} z} dz, \end{aligned}$$

changing variables. Following a parallel argument to [Muthen \(1990\)](#), it is possible to show that

$$\begin{aligned} E[x_i x_j] &= q E[y_i y_j | y_i > 0, y_j > 0] = \zeta \left. \frac{\partial^2 M(t)}{\partial t_i \partial t_j} \right|_{t=0} \\ &= \left(\frac{1}{\rho_0} + \frac{1}{\rho_\epsilon} \right) \left(\zeta r + \frac{1}{2\pi} \sqrt{1 - r^2} \right). \end{aligned}$$

¹⁸[Greene \(2012, chapter 19\)](#) derives the moments of censored Gaussian variables.

Thus we obtain

$$\begin{aligned} \text{Cov}(x_i, x_j) &= E[x_i x_j] - E[x_i] E[x_j] \\ &= \left(\frac{1}{\rho_0} + \frac{1}{\rho_\epsilon} \right) \left[\zeta r - \frac{1}{2\pi} (1 - \sqrt{1 - r^2}) \right]. \end{aligned} \quad (26)$$

The results in the text follow by applying Equations (24) and (26) and substituting for ζ using the following Lemma:

Lemma 1. *Let $y = (y_i, y_j)$ be distributed $y \sim \mathcal{N}(0, R)$, where R is defined in (25). Then*

$$\zeta = \Pr[y_i > 0, y_j > 0] = \frac{1}{4} + \frac{1}{2\pi} \text{ArcSin}(r).$$

Proof. Note that $z = y/\sigma$ has a bivariate standard normal distribution, with mean zero and covariance matrix $\begin{pmatrix} 1 & r \\ r & 1 \end{pmatrix}$. Let $\phi(z_1, z_2)$ be the joint density of this distribution. Since $\Pr[y > 0] = \Pr[y/\sigma > 0] = \Pr[z > 0]$, we can express q only in terms of $\phi(z_1, z_2)$. Using the obvious abuse of notation we can split this up into marginal and conditional density: $\phi(z_1, z_2) = \phi(z_1)\phi(z_2|z_1)$, where $\phi(z_1)$ is now just a univariate standard normal density. We now have

$$\begin{aligned} \zeta &= \int_0^\infty \int_0^\infty \phi(z_1, z_2) dz_2 dz_1 \\ &= \int_0^\infty \phi(z_1) \int_0^\infty \phi(z_2|z_1) dz_2 dz_1 = \int_0^\infty \phi(z_1) \Pr[z_2 > 0|z_1] dz_1. \end{aligned} \quad (27)$$

We know that $z_2|z_1 \sim N(rz_1, 1 - r^2)$, and so

$$\begin{aligned} \Pr[z_2 > 0|z_1] &= 1 - \Phi\left(\frac{-rz_1}{\sqrt{1 - r^2}}\right) \\ &= \Phi\left(\frac{rz_1}{\sqrt{1 - r^2}}\right), \end{aligned} \quad (28)$$

where Φ is a univariate standard normal CDF. Combining (27) and (28), we get

$$\zeta = \int_0^\infty \phi(z_1) \Phi\left(\frac{rz_1}{\sqrt{1 - r^2}}\right) dz_1.$$

The strategy is to first characterize $\frac{d\zeta}{dr}$ and then integrate to find ζ itself. Differentiating

under the integral,

$$\frac{d\zeta}{dr} = \frac{d}{dr} \left(\frac{r}{\sqrt{1-r^2}} \right) \times \int_0^\infty z_1 \phi(z_1) \phi \left(\frac{rz_1}{\sqrt{1-r^2}} \right) dz_1. \quad (29)$$

Consider the two factors in turn. For the first one,

$$\begin{aligned} \frac{d}{dr} \left(\frac{r}{\sqrt{1-r^2}} \right) &= \frac{d}{dr} \left[r (1-r^2)^{-1/2} \right] \\ &= (1-r^2)^{-1/2} + r \left(-\frac{1}{2} \right) (1-r^2)^{-3/2} (-2r) \\ &= (1-r^2)^{-1/2} + r^2 (1-r^2)^{-3/2} \\ &= \frac{1}{\sqrt{1-r^2}} \left(1 + \frac{r^2}{1-r^2} \right) \\ &= \frac{1}{\sqrt{1-r^2}} \left(\frac{1-r^2+r^2}{1-r^2} \right) \\ &= (1-r^2)^{-3/2}. \end{aligned}$$

For the second one, note that

$$\begin{aligned} \phi(z_1) \phi \left(\frac{rz_1}{\sqrt{1-r^2}} \right) &= \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z_1^2} \right] \times \left[\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{rz_1}{\sqrt{1-r^2}} \right)^2} \right] \\ &= \frac{1}{2\pi} e^{-\frac{z_1^2}{2} \left(1 + \frac{r^2}{1-r^2} \right)} \\ &= \frac{1}{2\pi} e^{-\frac{1}{2} \left(\frac{z_1}{\sqrt{1-r^2}} \right)^2} \\ &= \frac{1}{\sqrt{2\pi}} \phi \left(\frac{z_1}{\sqrt{1-r^2}} \right) \end{aligned}$$

Substituting the last two equations into (29), we have

$$\frac{d\zeta}{dr} = \frac{1}{1-r^2} \frac{1}{\sqrt{2\pi}} \int_0^\infty z_1 \frac{1}{\sqrt{1-r^2}} \phi \left(\frac{z_1}{\sqrt{1-r^2}} \right) dz_1.$$

The integral in this expression is equal to the expectation of a censored normal variable with mean zero and variance $1-r^2$ (explicitly, it is equal to $E[\max\{\tilde{z}, 0\}]$ where $\tilde{z} \sim N(0, 1-r^2)$).

This expectation (using Greene’s formula again) equals $\sqrt{1-r^2}\sqrt{\frac{1}{2\pi}}$. Thus we finally get

$$\begin{aligned}\frac{d\zeta}{dr} &= \frac{1}{1-r^2} \frac{1}{\sqrt{2\pi}} \times \sqrt{1-r^2} \sqrt{\frac{1}{2\pi}} \\ &= \frac{1}{2\pi} \frac{1}{\sqrt{1-r^2}}.\end{aligned}$$

Now we can solve for q explicitly using this differential equation and a boundary condition. For the boundary condition, note that if the two variables aren’t correlated, i.e. $r = 0$, then $\zeta = \frac{1}{4}$. We therefore have

$$\begin{aligned}\zeta(r') &= \frac{1}{4} + \int_0^{r'} \frac{d\zeta}{dr} dr \\ &= \frac{1}{4} + \frac{1}{2\pi} \int_0^{r'} \frac{1}{\sqrt{1-r^2}} dr\end{aligned}$$

The inverse sine function has the derivative $\frac{d\text{ArcSin}(r)}{dr} = \frac{1}{\sqrt{1-r^2}}$, and satisfies $\text{ArcSin}(0) = 0$, we obtain the required result:

$$\zeta(r) = \frac{1}{4} + \frac{1}{2\pi} \text{ArcSin}(r).$$

□

B Computational results

Section 2 derived the effects of news on volatility using an approximation. In this section, as a robustness check, we provide numerical evidence that the approximation we used yields correct results.

We make multiple draws of the asset’s fundamental value, $\theta \sim \mathcal{N}(100, 1)$. We call each of these draws a “market.” Preferences are CARA, as in Section 2. In each market, there is the realization of N signals of the form $s_i = \theta + \epsilon_i$, where $\epsilon \sim \mathcal{N}(0, \rho_\epsilon)$ with $\rho_\epsilon = 1$. We will consider integer values of N at (approximately) 20 equally spaced points in the interval $N \in [1, 40]$. There is a unit mass of consumers, where $\frac{1}{2}$ are rational/Bayesian and $\frac{1}{2}$ exhibit one of the behavioral biases described below. For each group, their prior beliefs follow the true distribution of θ , $\mathcal{N}(100, 1)$. We normalize the CARA coefficient and the rate of interest to be unity.

For each market, we compute the posterior beliefs for each group of traders, their demands, and the implied equilibrium price p . For each bias, and for each value of N , we make 50,000 draws of θ and then compute $\mathbb{V}[\theta/p]$ across these realizations of θ . This approach allows us to compute the volatility of returns $\mathbb{V}[\theta/p]$ directly, rather than use an approximation as in Section 2. For completeness, we present also regarding the average trading volume across markets, for different values of N .

Rational market. Here, we assume that all traders are rational/Bayesian.

Example 1: Overconfidence. Behavioral traders observe the true signals s_i , but update their beliefs as if signals had the precision $\hat{\rho}_\epsilon = 2\rho_\epsilon = 2$. This corresponds to small biases in the terminology of Section 2. The results for large biases are available on request.

Example 2: Stale news. Behavioral consumers observe “echoes” of the true signals. In particular, we assume that rational consumers observe N signals, while behavioral consumers observe each of these signals twice, for a total of $2N$ informative signals. The Figure below shows the case where the number of informative signals increases. The results when the number of repetitions increases are available on request.

Example 3: Conservatism. Behavioral investors overestimate the precision of their prior, using $\hat{\rho}_0 = 2\rho_0 = 2$.

Example 4: Inattention. We assume that behavioral traders can observe at most 20 signals. When $N > 20$, behavioral traders observe only 20 randomly selected signals from the set of all available signals. Effectively, all traders behave rationally for $N \leq 20$.

Example 5: Sentiments. We consider three sub-cases of sentiments. First, we consider deterministic optimism. Behavioral traders perceive the signal as $\hat{s}_i = s_i + b$, where s_i is the true signal (perceived by Bayesian traders) and b is a fixed scalar. We present the results for $b = 10$ (optimism). The results for $b = -10$ (pessimism) are similar and available upon request.

Second, we consider confirmation bias. We assume that, for signal realizations above the prior mean (100), behavioral these traders observe the true signal. However, for signals realizations below the prior mean, these traders interpret these signals to be exactly the same as their prior. That is, for $s_i \leq 100$, behavioral traders observe $\tilde{s}_i = 100$.

Finally, we consider “strong sentiments” which are stochastic. In each market, there is a draw of the level of “sentiment” experienced by behavioral traders. Sentiment is a random variable \tilde{b} , which is distributed $\mathcal{N}(0, V_b)$. We assume $V_b = 9$. In each market, behavioral consumers interpret each signal s_i to be $\tilde{s}_i = \theta + \tilde{b}$. Within each market θ , the realization \tilde{b} is the same for all behavioral traders and each signal is perceived with the same bias \tilde{b} (i.e., sentiment is positively correlated within a market). However, \tilde{b} is independent across markets. Notice that behavioral consumers are correct on average, since $\mathbb{E}[\tilde{b}] = 0$.

The figures below show the computational results. In all cases, they confirm the analytical (but approximate) comparative statics we obtained in Section 2.

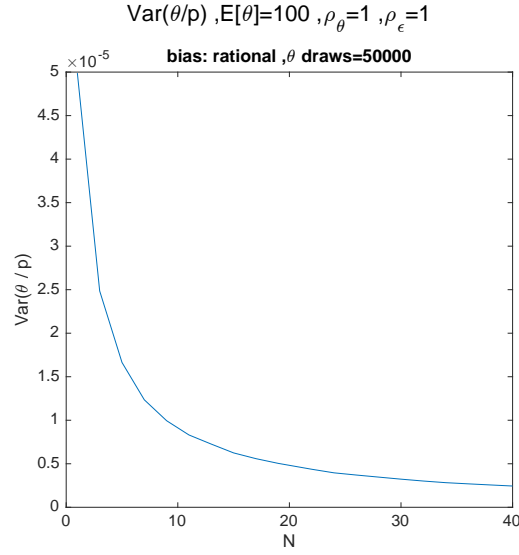


Figure 1: Volatility $\mathbb{V}[\theta/p]$ and trading volume $\mathbb{E}[T]$ across realizations of θ , for different values of N . In these panels, all traders are rational.

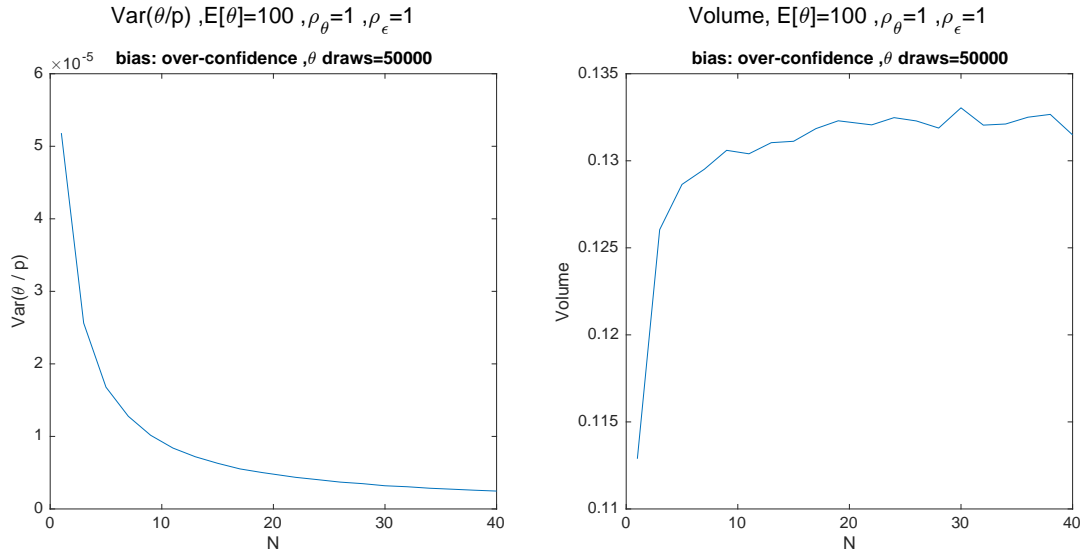


Figure 2: Volatility $\mathbb{V}[\theta/p]$ and trading volume $\mathbb{E}[T]$ across realizations of θ , for different values of N . In these panels, half of all traders exhibit over-confidence.

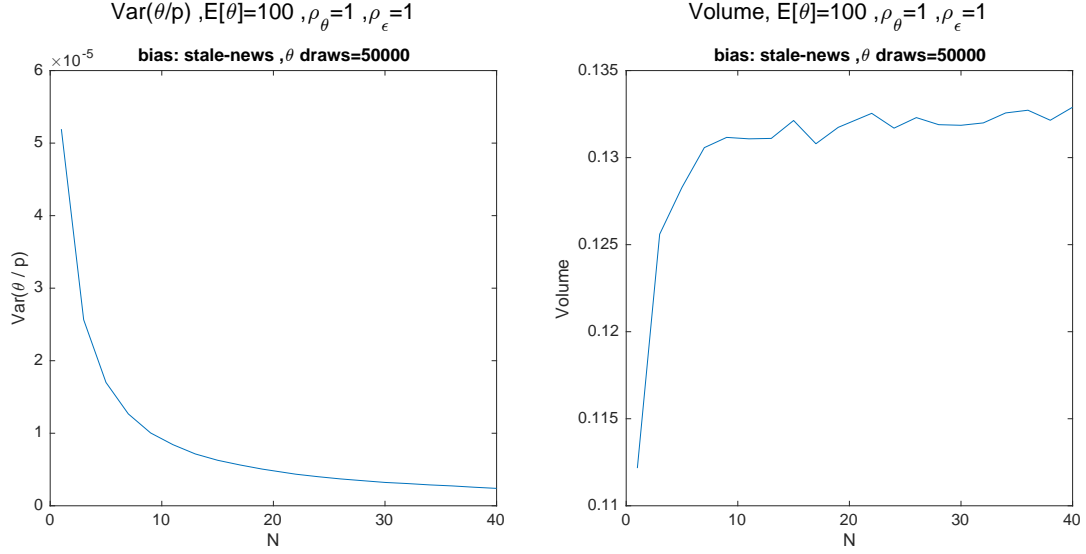


Figure 3: Volatility $\mathbb{V}[\theta/p]$ and trading volume $\mathbb{E}[T]$ across realizations of θ , for different values of N . In these panels, half of all traders exhibit stale news bias.

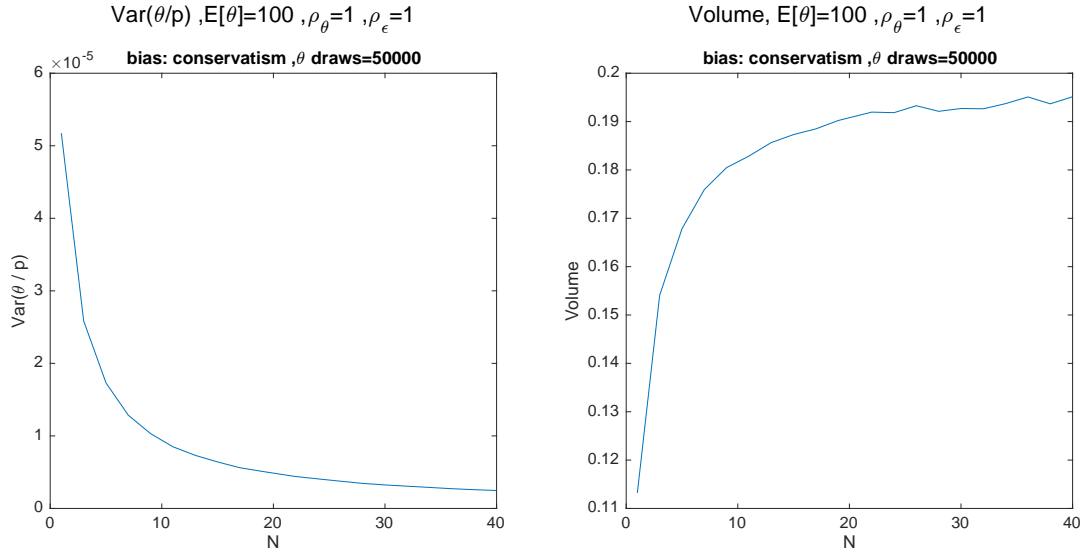


Figure 4: Volatility $\mathbb{V}[\theta/p]$ and trading volume $\mathbb{E}[T]$ across realizations of θ , for different values of N . In these panels, half of all traders exhibit conservatism.

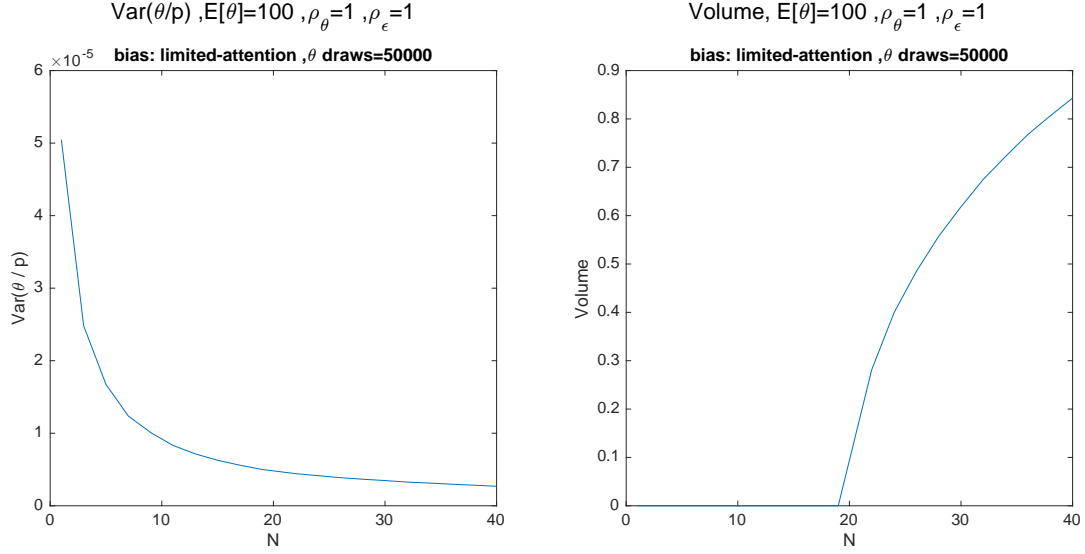


Figure 5: Volatility $\mathbb{V}[\theta/p]$ and trading volume $\mathbb{E}[T]$ across realizations of θ , for different values of N . In these panels, half of all traders exhibit limited attention.

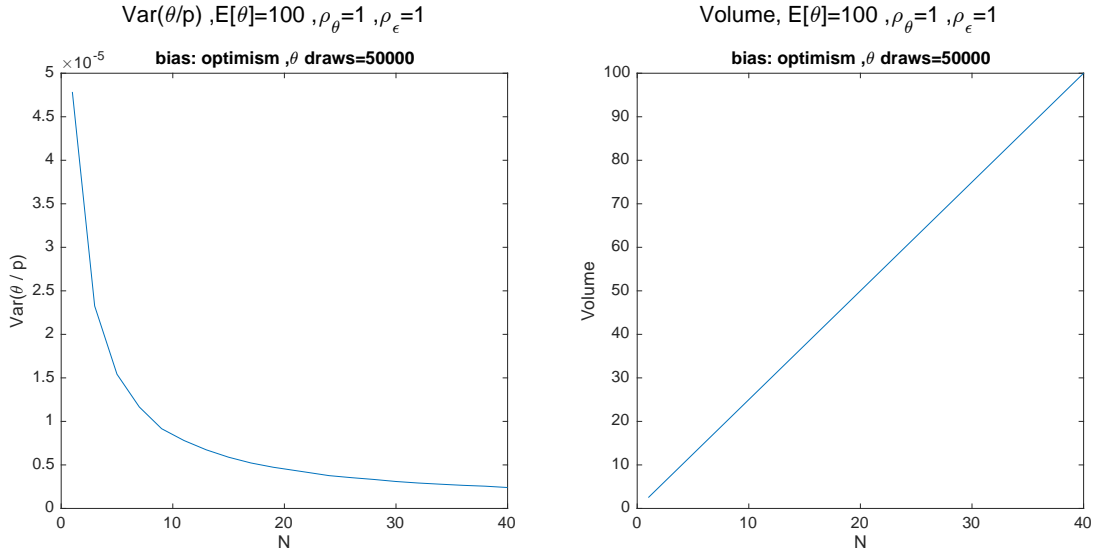


Figure 6: Volatility $\mathbb{V}[\theta/p]$ and trading volume $\mathbb{E}[T]$ across realizations of θ , for different values of N . In these panels, half of all traders exhibit optimism.

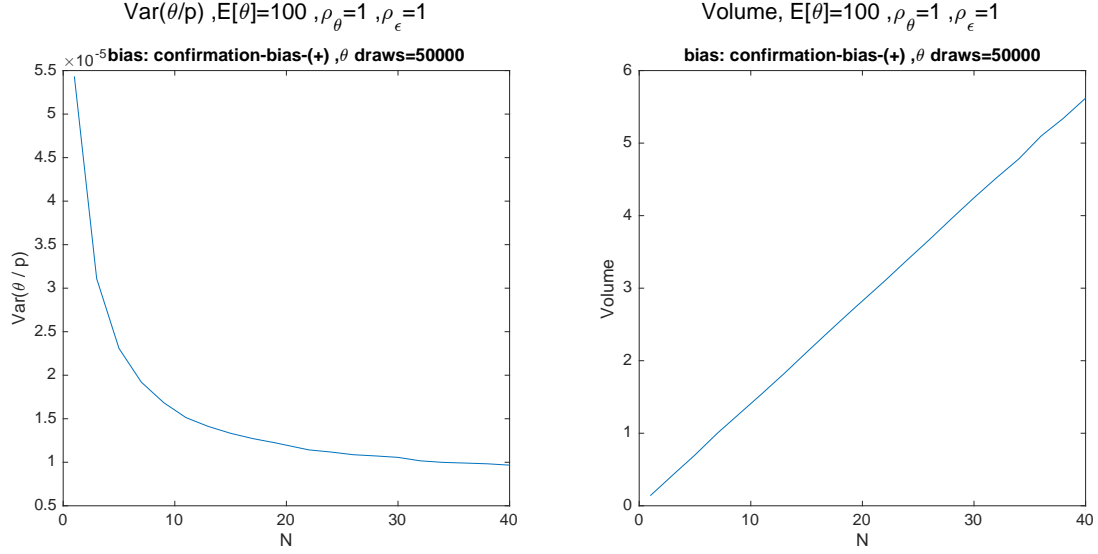


Figure 7: Volatility $\mathbb{V}[\theta/p]$ and trading volume $\mathbb{E}[T]$ across realizations of θ , for different values of N . In these panels, half of all traders exhibit confirmation bias.

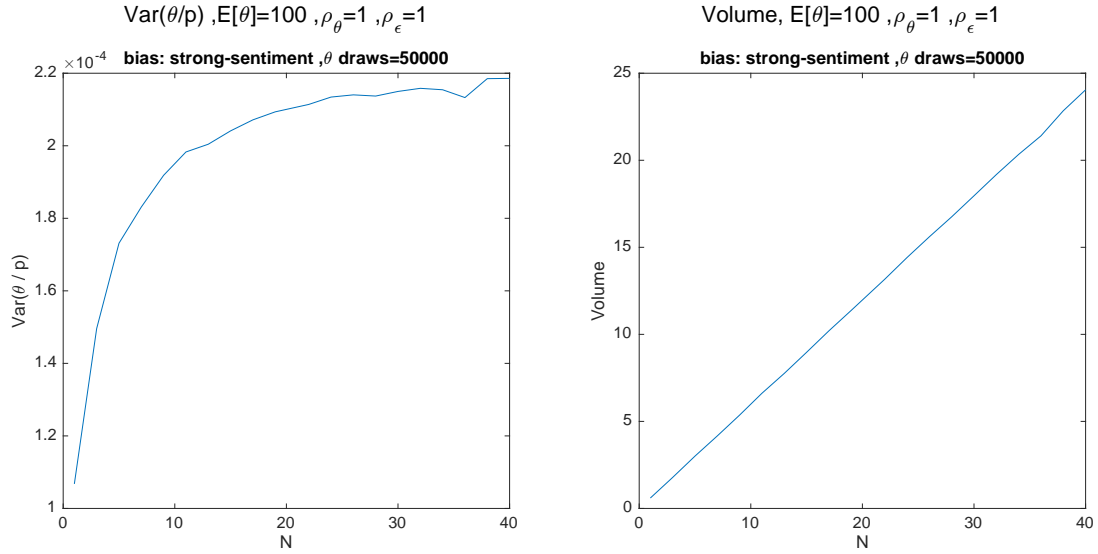


Figure 8: Volatility $\mathbb{V}[\theta/p]$ and trading volume $\mathbb{E}[T]$ across realizations of θ , for different values of N . In these panels, half of all traders exhibit strong sentiment.