ESSAYS IN FINANCIAL MARKETS

by

Xiaodi Zhang

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Business Administration

David Eccles School of Business

The University of Utah

August 2015

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Xiaodi Zhang**

has been approved by the following supervisory committee members:

Shmuel Baruch	, Chair	**04/15/2015** Date Approved
James Schallheim	, Member	**04/15/2015** Date Approved
Michael Cooper	, Member	**04/15/2015** Date Approved
Feng Zhang	, Member	**04/15/2015** Date Approved
Jingyi Zhu	, Member	**04/15/2015** Date Approved

and by **William Hesterly** , Chair/Dean of

the Department/College/School of **David Eccles School of Business**

and by David B. Kieda, Dean of The Graduate School.

ABSTRACT

This dissertation studies how heterogeneous opinions affect financial market outcomes, including price informativeness and trading volume. The dissertation contains two chapters. In both chapters, theoretical models are developed and then supportive empirical evidence is provided.

In the chapter "Index Trading and Its Effects on the Underlying Assets" (Chapter 1), I present a rational expectation model of index trading. The key finding is that the efficiency of each of the underlying stocks decreases as the proportion of index traders increases, while the efficiency of the index itself is unchanged. This result is achieved despite the fact that no arbitrage opportunities exist, i.e., the price of the basket (index) is the sum of its components. Using S&P 500 ETFs data, I show that the index contributes to price discovery in its underlying stocks. In addition, the regression analysis is consistent with the model predictions: index trading impairs efficiency of the component securities but does not have effects on the index itself.

In the chapter "News, Influence, and Evolution of Prices in Financial Markets" (Chapter 2), we study the influence of published views on the evolution of prices by constructing a theoretical model and using empirical work to test the model. Our sequential trade model demonstrates how the influence of published views creates patterns in prices and volume. Still, a "wisdom of the crowds" effect emerges endogenously in our framework and helps expunge such shared errors from the price, thus setting the paper apart from the information cascades literature. We use the timing of earnings announcements to test our model, and find evidence consistent with the theoretical predictions. The magnitude of the empirical effects is then used to calibrate the model, and our calibration exercise suggests that patterns in the evolution of prices are affected more strongly by the extent of the influence of the published view than by its accuracy.

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

I would like to express my deepest gratitude to my committee chair, Professor Shmuel Baruch, for his excellent guidance, caring and constant encouragement. His passion and creativity inspired me during my whole Ph.D. study. Without his persistent help, this dissertation would not have been possible.

I would like to thank my committee member, Professor James Schallheim, who provided important insights in my research and patiently corrected my writing. I would like to thank my other committee members, Professor Michael Cooper, Professor Feng Zhang, and Professor Jingyi Zhu, for their kind support and help. I would also like to thank Professor Peter Bossaerts and Professor Hendrik Bessembinder for their helpful comments on my dissertation.

In addition, a thank you to Professor Gideon Saar of Cornell University, who led me to the world of empirical market microstructure and motivated me as a role model.

I also owe my sincere appreciation to my colleagues at the University of Utah, especially to Huan Cai, Al Carrion, Barbara Chambers, Wenhao Yang, and Yan(Julian) Zhang.

Last but not least, I would like to thank my lovely wife, Aiting Gao, and my parents. They have been always standing by me, supporting me and encouraging me with their best wishes.

CHAPTER 1

INDEX TRADING AND ITS EFFECTS ON THE UNDERLYING ASSETS

1.1 Introduction

The index mutual funds and the index-based exchange-traded funds (ETFs) have become very attractive investment vehicles since their first introduction. By the end of the year 2013, index funds and index ETFs managed total net assets of about $1.7 trillion and $1.6 trillion, respectively.[1] By definition, the index fund is used to replicate the performance of a particular market index. Prior studies mostly focus on the U.S. stock index and the index future and examine which one is dominant in price discovery, for example, Kawaller, Koch, and Koch (1987), Stoll and Whaley (1990), and Chan (1992). These papers find that the futures market leads the cash market but present weaker evidence in the reverse direction. However, not many empirical studies examine the impact of index trading on its underlying assets. Yu (2005) investigates how the introduction of index ETFs affects the efficiency of individual component securities. Qin and Singal (2013) study the relation between passive ownership and price efficiency using a sample of S&P 500 and non-S&P 500 stocks. In this paper, I try to fill this gap by developing a theoretical model of index trading and providing empirical evidence for the model's predictions.

Traditionally, people consider the index fund a redundant asset since it can be easily replicated using its component securities. However, the environment of financial markets has changed since the introduction of index ETFs and the advent of high-frequency and algorithmic trading. Recent studies by Yu (2005) and Jovanovic and Menkveld (2012) show that the index related security delivers information about its components. Hasbrouck (2003) also finds that the S&P 500 ETF contributes information about its sector ETFs. These findings both convey a signal that index trading may contain some pieces of information on its underlying assets. When market participants with different views on asset values trade

[1] Investment Company FactBook 2014.

the index future or ETF, it becomes nonredundant. The existence of the index now affects prices of its underlying assets because these prices will adjust to a new equilibrium.

To study the effects of index trading on its component stocks, I investigate a rational expectation model with two types of investors: index traders and stock pickers. In my model, the indexers only trade the market portfolio (i.e., index fund or index ETF) but the stock pickers can pick any component stock to trade.[2] As in Bossaerts (1993), I do not model explicitly why indexers exist on the market. Investors might trade index because it is simple or because the execution of index incurs low costs. Given the large investment in index funds and index ETFs, it is a reasonable assumption that some of the investors on the market love to trade index.

More importantly, to capture the idea that different investors may possess different information sets, each investor in my model is endowed with personal views on the prospects of individual assets. Investors form their personal views via related news papers, TV programs, or the Internet. In the model, I assume that personal views are more likely to be correct and errors among these views are independent. This notion of views is similar to the imprecise information in Admati and Pfleiderer (1986) but here the formation of views requires no costs. My model set-up is different from models with private information, where information is precise and only a small number of investors can access it. In my model, everyone holds personal views which are self-generated signals (i.e., a piece of imprecise information) and the accuracy of views is very low. The view and the signal (or the piece of information) are mathematically equivalent so I will use these terms interchangeably in the paper.

In the equilibrium of my model, the price of market portfolio is equal to the sum of prices of individual stocks so there does not exist any arbitrage opportunity. The model predicts that individual stocks become less efficient when there are more index traders but the efficiency of the index itself stays constant. The efficiency in my model is defined as in Kyle (1985), which is the posterior variance of return. The reason for the above implications is that index traders do not utilize their views on individual assets. Instead, they use the combined views, i.e., views of the market portfolio. In the equilibrium, prices of individual stocks are not fully revealing since they only reflect indexers' views about the market as a whole. In my model, indexers want to cope with the winner's curse problem when they

[2]In this paper, I use the term "market portfolio" and "index fund or ETF" interchangeably. The term "index fund" is used here to be consistent with the empirical work.

make their investment decisions. Thus they optimally extract information from individual stock prices. As a result, the price of market portfolio is always fully revealing.

The model closely relates to prior theoretical work on the index and index-related securities. Subrahmanyam (1991) studies how the introduction of a security basket affects market liquidity and the informativeness of its underlying assets and finds that the final results depend on different model parameters. He also concludes that the basket security has no effects on the variance of price changes in individual securities when the number of informed traders is constant. Gorton and Pennacchi (1993) develop a model to study how liquidity traders choose between a single security and a basket. Both models have informed traders possessing accurate information and require the existence of liquidity traders. Thus the price informativeness of securities depends on where these liquidity traders trade and the model parameters. My model is different from theirs in two respects. First, every trader in my model has views although they are imprecise. Second, liquidity traders do not exist in my model and the inefficiency all comes from trading in the market. Higher index trading leads to higher posterior variances of individual stock returns and causes more information loss on the underlying assets.

I also examine the model predictions using real world data and in particular, I use S&P 500 ETFs as the index. These heavily traded ETFs provide me with a good setting to study information flow between the index and its underlying assets. Using the Hasbrouck (1995) information share method, I find that price changes in the index contribute to about 25.69% in the efficient price variance of component stocks. This result contradicts the traditional cognition that the index security is a simple basket of its underlying assets. This approximate one quarter of price discovery in the component stocks provides significant evidence that index trading contributes some relevant information.

In order to test the model's implications, I use realized volatility as an empirical proxy for the theoretical efficiency measure, posterior variance. I use realized volatility for the following reasons. First, it is an estimate of ex post variance. Second, it does not require specific distributional assumptions. Third, many prior studies examine the relation between realized volatility and trading volume to provide insights on how traders process new information. For example, Jones, Kaul, and Lipson (1994) and Chan and Fong (2006) present evidence that trading volume explains almost all of the asset's daily realized volatility. Since my model relates index trading to posterior variance, the realized volatility fits the model well and provides very good context to test the model's predictions. Following

Andersen, Bollerslev, Diebold, and Ebens (2001) (ABDE (2001) hereafter), I construct daily realized volatility by summing up squares of high-frequency intraday returns. Specifically, the 1-minute and 5-minute interval returns are used to construct the proxy.

For individual assets, I measure the proportion of index traders as follows: I estimate daily trading volume generated from index ETFs for each component stock and divide it by that stock's total daily trading volume. For the index, I sum all daily trading volume of its underlying stocks and divide its own daily volume by the achieved sum. Using a sample period from October 1, 2013 to December 31, 2013, I regress the daily realized volatility on the proxy of index traders for the index and each component stock. Consistent with the model implications, I find that proportion of indexers is positively and significantly related to realized volatility for component stocks but such relation does not exist in the index itself.

This paper also relates to the literature on price changes in the case of index additions and deletions. Most of the studies in this literature argue that index additions and deletions involve no information. Therefore the price changes caused by indexing are due to downward sloping demand curves. For example, Shleifer (1986) and Lynch and Mendenhall (1997) present evidence consistent with this hypothesis. Two recent studies provide evidence for other explanations. Denis, McConnell, Ovtchinnikov, and Yu (2003) find that additions to the S&P 500 index provide good information about companies' future prospects. Chen, Noronha, and Singal (2004) show that asymmetric price changes due to S&P 500 additions and deletions are caused by investors' awareness. Their finding is consistent with Merton (1987)'s model of market segmentation. My model together with its empirical evidence, however, sheds new light on this literature. Since index trading changes the information environment of its component stocks, index additions and deletions are no longer information free so prices will adjust to a new equilibrium. This may explain the price changes due to index additions and deletions to the extent that index trading alters prices of its component stocks.

My study contributes to the literature in at least two ways. First, I develop a rational expectation model to capture the idea that the index trading may contain relevant information about its underlying assets. In particular, index trading generates inefficiency in its components. Second, I empirically examine the impacts of index trading and present new evidence that it impairs the market quality of its underlying assets.

The remainder of the paper is organized as follows. Section 1.2 studies a model with heterogeneous agents and presents the model's prediction. Section 1.3 discusses the empirical methodology and describes the sample. Section 1.4 presents the empirical results of the tests, and Section 1.5 concludes the paper.

1.2 The Model

I consider a competitive economy with N risk averse traders, a fraction α of them are index fund traders and $1-\alpha$ are stock pickers. All agents have a continuous and differentiable concave utility function U. The economy contains three assets: a risk-free bond and two risky assets. The economy has one period and all agents consume at the end of the period. Each agent is endowed with $1/N$ share of the two risky assets' supplies and in addition, everyone has personal views on each asset's terminal value. Investors form their views by reading related newspapers, watching TV programs, or searching on websites or forums and their view formation incurs very low costs. In the model, I assume that each investor forms personal views with zero cost and also require that errors among views be independent. As I mentioned, each investor's view is equivalent to a piece of information and its accuracy is very low.

1.2.1 Assets

The risk-free bond pays 1 with certainty at the end of the period and serves as the numeraire in this economy.

The two risky assets are indexed by j, $j = 1, 2$. The price of risky asset j is denoted by P_j. The end-of-period value $\tilde{v}_j$ of risky asset j follows the binary distribution, equalling either v_h or v_l. I assume that the two risky assets are independent from each other and the supply for each is normalized to 1. The independent assumption is without loss of generality and simplifies the calculation. The prior distribution of each risky asset j is the following:

$$\begin{cases} p_{hj} = Prob(\tilde{v}_j = v_h) = 1/2 \\ p_{lj} = Prob(\tilde{v}_j = v_l) = 1/2. \end{cases}$$

In the economy, the market portfolio (or index fund) contains one share of risky asset 1 and one share of risky asset 2. That is to say, the terminal value $\tilde{V}_I$ of the market portfolio is given by the linear combination $\tilde{v}_1 + \tilde{v}_2$. So $\tilde{V}_I$ can take value $V_H = 2v_h$, $V_M = v_h + v_l$, or $V_L = 2v_l$. According to the prior distribution of risky assets and the independence assumption between them, the prior distribution of the index fund is:

$$\begin{cases} p_{HI} = Prob(\tilde{V}_I = V_H) = 1/4 \\ p_{MI} = Prob(\tilde{V}_I = V_M) = 1/2 \\ p_{LI} = Prob(\tilde{V}_I = V_L) = 1/4. \end{cases}$$

1.2.2 Personal Views

Every investor, in the model, is endowed with personal views on terminal values of both risky assets. I assume that the relation between a view and the terminal value of asset j is given by the linear function:

$$\tilde{v}_j^n = \tilde{v}_j(1 - \tilde{\epsilon}_j^n) + (v_h + v_l - \tilde{v}_j)\tilde{\epsilon}_j^n,$$

where $j = 1, 2$, n denotes the n^{th} trader, and $n = 1, 2, ...N$. The error term $\tilde{\epsilon}_j^n$ is a zero-one random variable, identically and independently distributed across all traders (both indexers and stock pickers) and the risky assets. With the above construction, a view of risky asset j also takes either v_h or v_l. In addition, I assume a trader's view is more likely to be correct than the prior belief implying that the error term equals zero with probability $\pi_j > 1/2$. In this paper, views are self-generated signals, which are costless and imprecise. Mathematically, a view and a signal (or a piece of information) are equivalent so I will use these two terms interchangeably.

Given views on individual assets, a view on the market portfolio then directly follows:

$$\tilde{V}_I^n = \tilde{v}_1^n + \tilde{v}_2^n,$$

where n denotes the n^{th} trader and V_I^n can also be one of the three values: V_H, V_M, or V_L.

1.2.3 Index Fund Traders

In my model, index fund traders or indexers are traders who love to trade the market portfolio. Therefore these traders will condition on their views of the market when making investment decisions. At the same time, these traders also want to cope with the winner's curse problem so they will optimally extract information from prices of individual assets. Then indexer i's information set includes his view of the market and prices of individual assets, i.e., $\{V_I^i = v_1^i + v_2^i, P_1, P_2\}$, where $i = 1, 2, ..., \alpha N$.

Each indexer maximizes his expected end-of-period utility function by allocating between the market portfolio and the risk-free bond conditioning on his information set. I denote indexer i's posterior distribution of the market as q_{XI}^i, where X corresponds to each possible value in the state space $\{V_H, V_M, V_L\}$ respectively. Then the i^{th} indexer's problem is written as:

$$\max_{\theta_{iI}} E[U(\tilde{W}_{i,1})|V_I^i = v_1^i + v_2^i, P_1, P_2] = \sum_X q_{XI}^i \cdot U(W_{i,1}^X) \tag{1.1}$$

subject to the budget constraint

$$\tilde{W}_{i,1} = (W_{i,0} - \theta_{iI}(P_1 + P_2)) + \theta_{iI}\tilde{V}_I, \tag{1.2}$$

where θ_{iI} is the demand for the index fund. Since the index fund comprises the asset 1 and 2 with share ratio 1 : 1, the indexer's demands for individual assets are $\theta_{i1} = \theta_{iI}$ and $\theta_{i2} = \theta_{iI}$.

1.2.4 Stock Pickers

Stock pickers allocate their wealth among the two risky assets and the risk-free bond to maximize their expected end-of-period utility. Stock pickers are indexed by s, $s = 1, 2, ...(1-\alpha)N$. Since these traders will choose the demand of two risky assets respectively, their state space e regarding the final values of assets is $\{(v_1 = v_h, v_2 = v_h), (v_1 = v_h, v_2 = v_l), (v_1 = v_l, v_2 = v_h), (v_1 = v_l, v_2 = v_l)\}$. I index each event in the state space by e. The information set of stock picker s contains his own views and prices of the two risky assets, i.e., $\{v_1^s, v_2^s, P_1, P_2\}$. I let the posterior probability distribution of each event for the stock picker s be q_e^s and then his problem can be written as:

$$\max_{\theta_{s1}, \theta_{s2}} E[U(\tilde{W}_{s,1})|v_1^s, v_2^s, P_1, P_2] = \sum_e q_e^s \cdot U(W_{s,1}^e) \tag{1.3}$$

subject to the budget constraint

$$\tilde{W}_{s,1} = (W_{s,0} - \sum_{j=1}^{2} \theta_{sj} P_j) + \sum_{j=1}^{2} \theta_{sj}\tilde{v}_j, \tag{1.4}$$

where θ_{sj} is the demand for asset j.

1.2.5 Equilibrium and Its Properties

A *rational expectation equilibrium* in this economy is defined as: (i) each indexer solves his maximization problem by allocating his wealth between the index fund and the risk-free bond, given his information set; (ii) each stock picker solves his full blown portfolio optimization problem, conditional on his information set; and (iii) prices P_1 and P_2 clear the relative markets.

The market clearing conditions are given by:

$$\begin{cases} \sum_{i=1}^{\alpha N} \theta_{i1} + \sum_{s=1}^{(1-\alpha)N} \theta_{s1} = 1 & \text{asset 1} \\ \sum_{i=1}^{\alpha N} \theta_{i2} + \sum_{s=1}^{(1-\alpha)N} \theta_{s2} = 1 & \text{asset 2.} \end{cases} \tag{1.5}$$

Market clearing conditions contain the demand from all traders, i.e., indexers and stock pickers.

The trading of the index fund plays a role when stock pickers make their investment decisions. This is because the demands for asset 1 and 2 from the indexers contain their views on the index portfolio. These views of the market will be reflected in the prices of individual assets through the market clearing conditions. In addition, views contain imprecise but relevant information on terminal values of assets so the equilibrium prices will be different from those in the model without personal views.

Since each trader's demands depend on his views of either individual assets or the market as a whole, equilibrium prices, formed from market clearing conditions (1.5), are functions of views from all traders. I denote $O^s = \{(v_1^1, v_2^1), (v_1^2, v_2^2), ..., (v_1^{(1-\alpha)N}, v_2^{(1-\alpha)N})\}$ the set of views of stock pickers and $O^i = \{V_I^1, V_I^2, ..., V_I^{\alpha N}\}$ the set of views of indexers. Then the view set of all traders is given by $O = \{O^s, O^i\}$ and the prices can be written as $P_1(O)$ and $P_2(O)$.

Given the above assumption about assets, I need the following lemmas to find the economy's equilibrium prices.

Lemma 1 *Let X be a discrete random variable taking values $\{0, 1, 2, ..., k\}$. The unknown probability distribution of X is the vector $\mathbf{p} = \{p_0, p_1, p_2, ..., p_k\}$, where $p_i = Prob(X = i)$ and $\sum_{i=0}^{k} p_i = 1$. In addition, $I(X = i)$ is an indicator function, equalling 1 if $X = i$ and 0 otherwise. Let $\{X_1, X_2, ..., X_N\}$ be a sample of N independent such random variables. Let $N_i = \sum_{n=1}^{N} I(X_n = i)$, where $\sum_{i=0}^{k} N_i = N$. Then $\hat{\mathbf{p}} = \{N_0, N_1, N_2, ..., N_k\}$ is sufficient for $\mathbf{p}$.*

Proof. See Appendix A.2.

Using this lemma, I can find sufficient statistics for O, given v_1 and v_2, for indexers and stock pickers, respectively. Since indexers only care about the value of the index, in their view, the information set O^s of stock pickers is equivalent to $O^{s\prime} = \{V_I^1, V_I^2, ..., V_I^{(1-\alpha)N}\}$. For this reason, the total information set O is reduced to $O' = \{O^{s\prime}, O^i\}$ for indexers. So by Lemma 1, their sufficient statistic for O' given v_1 and v_2 is $N^i = \{N_{Hs}, N_{Ms}, N_{Ls}, N_{Hi}, N_{Mi}, N_{Li}\}$, where N_{Hs}, N_{Ms}, and N_{Ls} correspond to the numbers of occurrence of V_H, V_M, and V_L in $O^{s\prime}$ and N_{Hi}, N_{Mi}, and N_{Li} are the relative numbers in O^i. Stock pickers, on the other hand, select their demands for asset 1 and asset 2 individually so they try to extract information on each risky asset from O. Then according to Lemma 1, for stock pickers, their sufficient statistic for O conditional on v_1 and v_2 is given by $N^s = \{N_{h1}, N_{l1}, N_{h2}, N_{l2}, N_{Hi}, N_{Mi}, N_{Li}\}$, where N_{h1} and N_{l1} are the numbers of occurrence of v_h and v_l for asset 1 in O^s, and N_{h2} and N_{l2} denote the corresponding numbers for asset 2 in O^s. I combine N^i with N^s to form a sufficient statistic $N^o = \{N_{h1}, N_{l1}, N_{h2}, N_{l2}, N_{Hs}, N_{Ms}, N_{Ls}, N_{Hi}, N_{Mi}, N_{Li}\}$ of the information set O for all traders. Then, for indexers, N^o is equivalent to N^i which is a sufficient statistic for the conditional joint probability mass function $h(O'|V_I = v_1 + v_2)$; stock pickers consider N^o as N^s which is a sufficient statistic for $h(O|v_1, v_2)$.

Together with the above analysis, I also need another lemma to solve the equilibrium problem. Before describing the next lemma, I define the conditional probabilities for the terminal values of risky assets and the index fund for later use. I let π_{yxj} be the conditional probability of a view of asset j on the value of asset j, where $j = 1, 2$, $y = \{v_h, v_l\}$ is the value set of views, and $x = \{v_h, v_l\}$ is the value set of risky assets. I denote π_{YXI} the corresponding conditional probability for values of the index, where Y and X are value sets of views and the index, respectively, both equalling $\{V_H, V_M, V_L\}$. I list all relevant variables in Table 1.1.

Lemma 2 *The conditional joint probability mass function $Prob(V_I^i, O|V_I = v_1 + v_2)$ for indexers is equal to $Prob(O'|V_I = v_1 + v_2)$, where $O' = \{O^{s\prime}, O^i\}$; and the conditional joint probability mass function $Prob(v_1^s, v_2^s, O|v_1, v_2)$ for stock pickers is equal to $Prob(O|v_1, v_2)$.*

Proof. See Appendix A.2.

Lemma 2 simply says that for each trader, if he knows the information contained in O(or O'), his individual opinion provides no additional information regarding v_1 and v_2. In other words, all index traders have the same posterior belief of the market portfolio and all stock pickers have the same belief regarding individual assets.

With the above two lemmas, I can characterize the equilibrium prices $P_1^*(O)$ and $P_2^*(O)$. By Lemma 1 and its subsequent analysis, given the realization of random variables, the information conveyed in the equilibrium price pair $\{P_1^*(O), P_2^*(O)\}$ is equivalent to N^o. So the individual investor's information set is reduced to $\{V_I^i, N^o\}$ or $\{v_1^s, v_2^s, N^o\}$. Using Lemma 2, this is further reduced to N^i or N^s. With the relevant information in mind, each investor can calculate his posteriors regarding the terminal values. Then the equilibrium price pair is achieved by solving (1.1), (1.3), and (1.5). So by looking at the prices of two risky assets, investors can infer N^o and then by conditioning on this inferred information, they form their demands such that the corresponding prices clear the asset markets. In this sense, the price pair achieved above constitutes the rational expectation equilibrium.

After characterizing how to solve the equilibrium prices, I examine the model's implications. I have the following proposition for the model with personal views:

Proposition 1 *The posterior variance of return for each risky asset j in the index increases with the proportion of index traders α; the posterior variance of the index return, however, is independent of α.*

The above proposition actually considers how trading in the index affects the price generation of its underlying assets. Traditionally, investors take the index fund as a redundant asset, i.e., a simple basket of underlying components which provides no additional information. However, the development of index ETF and trading technologies has changed the equity market significantly. Hasbrouck (2003) finds that the S&P 500 ETF contributes to its sector ETFs' price discovery more than the other way round. Yu (2005) and Jovanovic and Menkveld (2012) also show that price changes of index related securities convey information about their components. My model shows that index trading makes the prices of its component assets less informative or efficient in terms of the posterior variance of their returns. The inefficiency arises since indexers do not utilize their views on individual assets. Instead, index traders trade on their views of the index portfolio. Therefore stock pickers cannot tell exactly the individual opinion for each asset in the cases of $V_I^i = V_M$. For this reason, they have to consider all possible combinations of v_1^i and v_2^i that give $V_I^i = V_M$ and this makes the underlying assets less informative.

The system of (1.1), (1.3), and (1.5) produces two complicated functions thus making the model less tractable therefore I resort to numerical solutions. Figure 1.1 gives a numerical example of the relations described in Proposition 1. In this example, I assume that each investor has a quadratic utility function:

$$U(\tilde{W}) = \tilde{W} - \frac{1}{2b}\tilde{W}^2 = -\frac{1}{2b}(\tilde{W} - b)^2 + \frac{b}{2}, \tag{1.6}$$

where b is large enough to avoid the problem of negative marginal utility. In addition, I let $N = 100$, $v_h = 11$, and $v_l = 10$. I also assume $\pi_1 = \pi_2$ for simplification and take the error precision as 0.60. I simulate the model's random variables based on different αs (i.e., proportion of index traders). With the specific realization of random variables and the utility in (1.6), there always exists a unique equilibrium price pair such that $v_l \le P_1^* \le v_h$ and $v_l \le P_2^* \le v_h$. Then the investor can always establishes a one-to-one correspondence between $\{P_1^*, P_2^*\}$ and the sufficient statistic N^o.

I calculate the equilibrium prices of two risky assets and the corresponding posterior variances of returns, and then plot posterior variances against different αs in Figure 1.1. Figure 1.1 shows clearly that the posterior variance of each asset j increases with α but the posterior variance of the index stays constant.

My model is different from models in Subrahmanyam (1991) and Gorton and Pennacchi (1993) in that I do not have liquidity traders but instead have index traders. In their models, there exist some liquidity traders and it is these traders who provide camouflage for the informed traders who possess very accurate information. The price informativeness depends on where the liquidity traders trade. Since the introduction of a basket security may change the investment decisions of liquidity traders, the price informativeness may change correspondingly. However, my model does not require liquidity traders but instead assumes existence of indexers with views. This nonfully revealing equilibrium due to index trading differentiates my model from the traditional noisy rational expectation equilibrium which requires the existence of liquidity traders, noisy supply, or nontradable endowment shock. It is the trading of indexers (i.e., the constrained trading in the index) that generates the information loss in prices of component stocks. The more indexers exist on the market, the higher posterior variance each component stock has. This prediction is different from the conclusion in Subrahmanyam (1991) that the basket security has no effects on the variability of price changes of its component securities.

1.3 Empirical Methodology

My model allows me to study how index fund trading affects its underlying assets using real world data. The component stock becomes less efficient in terms of its posterior variance when there are more indexers trading on the market. Empirically, I use realized volatility to measure the posterior variance. The realized volatility is an estimate of ex post variance, which is just the concept used in my model. Furthermore, unlike the stochastic or implied volatility, it does not depend on restrictive distributional assumptions. In addition, prior literature studies how market participants react and process new information by examining the relation between volatility and trading volume and in those studies, realized volatility is widely used. For the above reasons, I consider realized volatility a good fit of my model.

In the test, I use S&P 500 ETF as the index since it is the most widely replicated index ETF on the market. This highly traded ETF also enables me to study the information flow between the index and its underlying assets by applying the information share approach in Hasbrouck (1995). The information share method can provide empirical evidence on whether the index contains some information about its component assets in short time horizons.

1.3.1 Hasbrouck Information Share

The information share method in Hasbrouck (1995) is used to measure the contribution to price discovery from each market in two or more related markets. It assumes that there is a single random walk component, called the efficient price, which is common to prices from all markets. But this method can also be extended to include prices which may not cointegrate with each other. Within the context of studying information contribution between index and the underlying assets, there no longer exists a single random walk common to all prices. Specifically, I construct the following multivariate price vector:

$$P = [Midquote\_Index_t \quad Midquote\_Stock_{j,t}]' \tag{1.7}$$

for each component j in the index, where $Midquote\_Index_t$ represents the midquote (ask plus bid divided by two) of the index at time t and $Midquote\_Stock_{j,t}$ denotes the midquote of the stock j at time t. Then the vector error correction model (VECM) is written as:

$$\Delta P_t = B_1 \Delta P_{t-1} + B_2 \Delta P_{t-2} + ... + B_M \Delta P_{t-M} + \gamma(z_{t-1} - \mu) + \epsilon_t, \tag{1.8}$$

where ϵ_t is a 2×1 vector of zero-mean innovations with variance matrix Ω, B_i is a 2×2 matrix of autoregressive coefficients corresponding to lag i of the price changes, and $\gamma(z_{t-1} - \mu)$ is the error correction term with $\mu = E(z_t)$. The VECM also has a moving average representation:

$$\Delta P_t = \epsilon_t + A_1 \epsilon_{t-1} + A_2 \epsilon_{t-2} + \cdots, \tag{1.9}$$

where A_i matrices are moving average coefficients and the sum of all the moving average coefficients is denoted as $A(1) = I + A_1 + A_2 + \cdots$.

With the above systems, I can find the proportion of stock j's innovation variance that is attributed to the innovation of the index and also the other way round. Since $A(1)$ accounts for the permanent impact innovations, then the total variance of permanent price changes for the l^{th} component in P is $\Sigma_{ll} = [A(1)\Omega A(1)']_{ll}$, where Σ is the covariance matrix of the price innovations and Σ_{ll} is the l^{th} diagonal element. If Ω is diagonal, the information contribution of the k^{th} component in P to the l^{th} security is given by:

$$S_l^k = \frac{a_{lk}^2 \Omega_{kk}}{\Sigma_{ll}^2}, \tag{1.10}$$

where a_{lk} is the k^{th} element in the l^{th} row of the sum matrix $A(1)$ and Ω_{kk} is the k^{th} diagonal element in Ω. If Ω is not diagonal, the information share is not exactly identified. In this case, using the Cholesky factorization of Ω one can determine the upper and lower bounds of the information share.

1.3.2 Index Trading and the Realized Volatility

The main conclusions of my model are in Proposition 1 which says that the higher proportion of index traders a component stock contains, the higher posterior variance it has but the efficiency of the index itself is independent of such a proportion. In order to test these predictions empirically, I first develop a variable that reflects the proportion of index trading for a component stock. I define my main variable as follows:

$$Index\_Trading\_Portion_{j,t} = \frac{Indextrading\_Dvol_{j,t}}{Trading\_Dvol_{j,t}}, \tag{1.11}$$

where $Indextrading\_Dvol_{j,t}$ is an estimate of the dollar volume generated by index trading for stock j on day t and $Trading\_Dvol_{j,t}$ is the total dollar volume for stock j on day t. For the index itself, the corresponding equation is written as:

$$Index\_Trading\_Portion_{I,t} = Index\_Dvol_{I,t} / \sum_j Trading\_Dvol_{j,t}, \tag{1.12}$$

where $Index\_Dvol_t$ is the dollar volume of all ETFs and $\sum_j Trading\_Dvol_{j,t}$ is calculated by summing all the dollar volume of its component stocks on the trading day t.

I then follow ABDE (2001) to construct the empirical proxy for posterior variance: realized volatility. ABDE (2001) shows that, under weak regularity conditions, the summed square of return within infinite small interval converges almost surely to the integrated latent volatility. Thus, by summing sufficiently finely sampled high-frequency returns, an arbitrarily accurate measure of daily return volatility can be constructed. A practical estimator of the integrated volatility is then computed by summing up all the intraday squared returns over many small intervals within a day. This estimator is the realized volatility and defined as:

$$\sigma^2_{j,t}(m) = \sum_{k=1,..,m} r^2_{j,t+k}, \tag{1.13}$$

where $\sigma_{j,t}(m)$ is the realized volatility for stock j on day t by dividing the day into m intervals and $r_{j,t+k}$ is the natural logarithmic return of the k^{th} interval. This approach is in line with the previous work by French, Schwert, and Stambaugh (1987) and Schwert (1989) who calculate the monthly realized stock volatilities using daily returns. The realized volatility is obtained for each component stock on every trading day. I also construct the realized volatility $\sigma_{I,t}(m)$ for the index itself using the same method.

ABDE (2001) also mentions that, practically, it is impossible to construct an estimator free of measurement error due to the bid-ask bounce and the uneven spacing of the observed prices. Since I use the S&P 500 ETF as the index, its component stocks are highly liquid so nonsynchronous trading is not a problem. In order to reduce the noise generated by the bid-ask bounce, I use midquotes to calculate interval returns.

With the above proxies, I can examine the relation between proportion of index traders and posterior variance. I regress the realized volatility on $Index\_Trading\_Portion$ for the index and each component stock, controlling for daily dollar trading volume. The control variable is in line with Jones, Kaul, and Lipson (1994) and Chan and Fong (2006) who find that trading volume explains almost all of the daily volatility. Specifically, I evaluate effects of the index trading by analysing the following equation:

$$\sigma_{j,t}(m) = \beta_0 + \beta_1 Index\_Trading\_Portion_{j,t} + \beta_2 log(Trading\_Dvol_{j,t}) + \epsilon_{j,t}, \tag{1.14}$$

where $Trading\_Dvol_{j,t}$ is the daily dollar trading volume for stock j on day t, $\sigma_{j,t}(m)$ is defined in equation (1.13), $Index\_Trading\_Portion_{j,t}$ is defined in (1.11), and $\epsilon_{j,t}$ is the error term. These variables are discussed in more detail in section 1.3.3. The above equation is estimated for each individual stock using daily observations. I also evaluate the corresponding equation for the index itself:

$$\sigma_{I,t}(m) = \beta_0 + \beta_1 Index\_Trading\_Portion_{I,t} + \beta_2 log(Index\_Dvol_{I,t}) + \epsilon_{I,t}. \tag{1.15}$$

where, in this case, volatility is the index's realized volatility $\sigma_{I,t}(m)$ and daily dollar volume is the index's trading volume $Index\_Dvol_{I,t}$.

1.3.3 Sample and Variable Construction

The primary data sources used in this study are TAQ and CRSP. Following Hasbrouck (2003), I consider the sample period from October 1, 2013 to December 31, 2013, the most recent three-month period during which TAQ data are available on WRDS when I started the paper's empirical study.

The index I use in this study is the S&P 500 value-weighted index ETF. Although my model applies to any index fund, there are several advantages of using this ETF. First, the S&P 500 index is one of the most commonly followed equity indices. At the end of 2013, mutual funds indexed to the S&P 500 alone account for 33% of all assets invested in the index funds (excluding ETFs).[3] Second, it contains a very diverse set of companies and is a good representation of the U.S. equity market. Third, the S&P 500 index ETF is heavily traded. On April 9, 2013, the average daily volume of the SPDR (an S&P 500 value-weighted ETF) itself was 117 million shares, the highest volume among all ETFs.[4] This highly liquid market provides me an ideal context to study the information transmission between the index and its underlying assets. In this paper, I only include three value-weighted ETFs in the sample: SPDR S&P 500 (SPY), iShares core S&P 500 (IVV), and Vanguard S&P 500 (VOO).[5]

During the sample period, there are several changes to the S&P 500 index. I list the corresponding companies and reasons for additions and deletions in Table 1.2.[6] So after adding those companies already deleted from the index, in total I have 507 stocks for the sample period.

I construct all prices used in the empirical study from TAQ database. Specifically, on each day, I define the opening price as the midquote for the quote in TAQ with MODE=10 from the stock's primary listing market or the last regular midquote before 9:31 am if such a quote does not exist.[7] I define the closing price as the midquote for the quote in TAQ with

---

[3] Investment Company FactBook 2014

[4] See Wiki: http://en.wikipedia.org/wiki/SPDR_S%26P_500_Trust_ETF

[5] I exclude the equal-weighted ETF, e.g., Guggenheim S&P 500, since the S&P 500 index is constructed based on market capitalization.

[6] See Wiki: http://en.wikipedia.org/wiki/List_of_S%26P_500_companies

[7] Following Boehmer, Saar, and Yu (2005), I apply certain criteria to screen the TAQ data. I keep quotes with TAQ's mode equal to 0, 1, 2, 3, 6, 10, 12, 23, 24, 25, or 26. I eliminate quotes with nonpositive ask or bid prices, or where the bid price is higher than the ask. I keep trades for which TAQ's CORR field is equal

MODE=3 from the stock's primary listing market or the midquote prevailing at 4:00 pm if such a quote does not exist. In addition to the opening and closing prices, I also define the 1-minute (5-minute) midquote as the last regular midquote in every minute (5 minutes) from 9:30 am to 4:00 pm in each day. The 1-minute midquotes are used in estimating the Hasbrouck information share.

I compute two realized volatilities using 1-minute and 5-minute midquotes, respectively. Following the literature, I calculate the natural logarithmic return for each 1-minute or 5-minute interval and sum up all these squared interval returns to get the daily volatility measures $\sigma_{j,t}(m)$ for each stock j and day t. When constructing volatilities $\sigma_{I,t}(m)$ for the index, I only use midquotes of SPY because this ETF is the most heavily traded among the three and often considered as a proxy for the market.[8]

To calculate my main variable $Index\_Trading\_Portion$, I need to estimate the index trading $Indextrading\_Dvol$ and the total trading $Trading\_Dvol$ for stock j on day t. For each day, I sum all dollar volume from three ETFs in TAQ as the $Index\_Dvol_{I,t}$ and also construct $Trading\_Dvol_{j,t}$ for each stock and day from TAQ. Since the S&P 500 index follows the value-weighted rule, I estimate volume generated by the index trading as:

$$Indextrading\_Dvol_{j,t} = \underbrace{\frac{Price_{j,t} \times Share\_Outstanding_{j,t}}{\sum Price_{j,t} \times Share\_Outstanding_{j,t}}}_{value-weight\ of\ stock\ j\ on\ day\ t} \times \underbrace{Index\_Dvol_{I,t}}_{daily\ volume\ of\ ETFs},$$

where $Price_{j,t}$ is the average of the opening and closing prices and $Share\_Outstanding_{j,t}$ is the shares outstanding for the stock which is from CRSP. Then for each stock and each day, I calculate an $Index\_Trading\_Portion$ as in equation (1.11) and winsorize this number at the top and bottom 0.5%.[9] To calculate this proportion for the index, I divide $Index\_Dvol_{I,t}$ by the total dollar volume of its component stocks on each trading day as in equation (1.12). I also construct the market capitalization $MktCap$, which equals the product of price and shares outstanding, for each stock and day. After constructing the above variables, I have in total 32195 stock-day observations and 64 observations for the index. During the sample

to either zero or one, and for which the COND field is either blank or equal to @, B, J, K, S, E, or F. I also exclude trades with nonpositive prices.

[8]SPY is heavily traded on multiple exchanges so the midquotes of SPY are constructed by using the National Best Bid and Offer (NBBO).

[9]This construction certainly underestimates the index trading since I exclude the equal-weighted and levered ETFs and do not consider the volume from traditional index funds.

period, DELL's bid and ask quotes do not change too much so I drop this stock when conducting related tests. Thus I have in total 506 stocks and 32181 stock-day observations left. The summary statistics of relevant variables for all component stocks and the index are provided in Table 1.3 and Table 1.5, respectively.

1.4 Empirical Results

1.4.1 Hasbrouck Information Share

Before analysing the empirical results of my model prediction, I first discuss results of the Hasbrouck information share. In estimating the VECM, I use the 1-minute midquotes, thirty lags and include thirty moving average coefficients in the sum matrix A. For each component stock j in the S&P 500 index, I compute an estimate of the information share attributed to the price changes in the index for the whole three months by only including prices of the index ETFs and stock j in (1.7). I use the midquotes of SPY in the price vector (1.7) because this ETF is the most heavily traded among the three and often considered as a proxy for the market. This information share is calculated by assigning the index ETF precedence (i.e., placing the index first in the price vector P in (1.7)) so the estimate is approximately the maximal information contribution from the index.

Table 1.6 provides the summary statistics on price discovery from SPY to its 500 components. It shows that the mean and median information contribution from SPY to the underlying assets is 25.69% and 24.61%, respectively. This around one quarter of the price discovery delivers reasonable evidence that the index provides material information to its component stocks, at least in short time horizons. In addition, I rotate the price vector and put the stock midquote first in P to estimate the maximal information share from stocks to the index. Table 1.7 presents the results and shows that the mean and median are 21.57% and 20.31%, respectively. This slightly lower information contribution from stocks to the index is somewhat surprising because traditionally investors would think there is more information flow from stocks to the index but not the other way round.

Yu (2005) and Jovanovic and Menkveld (2012) also find that the index related securities contribute to price discovery of individual assets. They mention that this is because price changes in the market index provide valuable hard information. In addition, Hasbrouck (2003) studies information flow between the S&P 500 ETF and its sector ETFs. He finds that the index has larger information contribution to its sector ETFs and his results are consistent with relatively low production of information at the sector level. With my findings in Table 1.6 and Table 1.7 and also the results from prior literature, it is reasonable to state that the index trading contains material information about its underlying assets.

1.4.2 Realized Volatility

Does index trading impair the market quality of its underlying assets? In this subsection, I address this question by examining the relation between proportion of index trading and realized volatility. Specifically I evaluate the main regression (1.14) for each component stock which controls daily dollar trading volume and present the regression results in Table 1.8. The coefficient in Table 1.8 is the average of coefficients from all the individual regressions and t-statistic and Adjusted R^2 are computed in the same way. t-statistics are reported in parentheses.

Using realized volatility as the dependent variable, I find a positive relation between proportion of index trading and volatility for the component stocks. The results in Table 1.8 are statistically significant for both 1-minute and 5-minute volatility measures. This finding is consistent with the prediction in Proposition 1 and the pattern shown in Figure 1.1 for the underlying assets. Moreover, I also find that dollar trading volume is significant in explaining the stock's daily volatility and this is consistent with prior literature.

In my model, the two component assets are identical by assumption so the model produces no size effect. In reality, however, firms are different in their capitalization. Therefore it is also interesting to see if size plays a role in the index trading. Thus I divide the 506 (excluding DELL) stocks into three size groups: large, medium, and small, based on their market capitalization at the beginning of the sample period. I provide the summary statistics of relevant variables within each size group in Table 1.4. The regression results are provided in Table 1.9. The empirical results within each size group are qualitatively similar to those in Table 1.8, which are calculated by using the total 506 stocks. The effect of index trading is stronger in the large size group and it is confirmed by the larger proportion of index traders in this category.

I also evaluate equation (1.15) for the S&P 500 ETFs and the results are presented in Table 1.10. Consistent with my model prediction, the proportion of index traders has no effects on volatilities of the index.

1.5 Conclusions

Index funds and index ETFs are both very popular investment vehicles among investors. These days, index ETFs have grown particularly quickly and they attract nearly twice the flows of traditional index funds since 2007 in the U.S. market.[10] By definition, the index fund is redundant because it is easily replicated by the underlying securities. However, with the development of trading technology, index funds, especially index ETFs, may attract investors with different opinions on assets' payoffs. Therefore the index may contain valuable pieces of information from these heterogeneous traders. As a consequence, the prices of its component securities will adjust to a new equilibrium to reflect these valuable opinions conveyed by the index trading.

The purpose of this paper is to study the influence of index trading on its component stocks. I first develop a rational expectation model to capture the idea that the index contains valuable opinions on the assets' terminal values. In equilibrium, the index's price is equal to the sum of prices of its component stocks so there is no arbitrage opportunity. The model predicts that the efficiency of the underlying assets decreases with proportion of indexers but the efficiency of the index itself stays unchanged, where efficiency is defined as the posterior variance of return. Unlike the models with private information, such as Subrahmanyam (1991) and Gorton and Pennacchi (1993), the inefficiency all comes from the index traders because they only trade the index that follows a predetermined rule, for example, the value-weighted rule.

Then I test my model using real world data. Before examining the model implications, I confirm that the index does contribute to price discovery in its component stocks, at least in short horizons, using the S&P 500 ETF as the index. I use the realized volatility as a proxy for the posterior variance, the theoretical measure of price efficiency. I show that index trading is positively and significantly related to the volatility measures. On the other hand, such results do not exist in the index ETF. This differs from prior works which focus on lead-lag effects between the index future and the stock index. My empirical findings also shed new light on the studies of price effects when there are index additions or deletions. Overall, the empirical findings are consistent with my model predictions that index trading makes its underlying assets less efficient but does not affect the index itself.

[10] Investment Company FactBook 2014

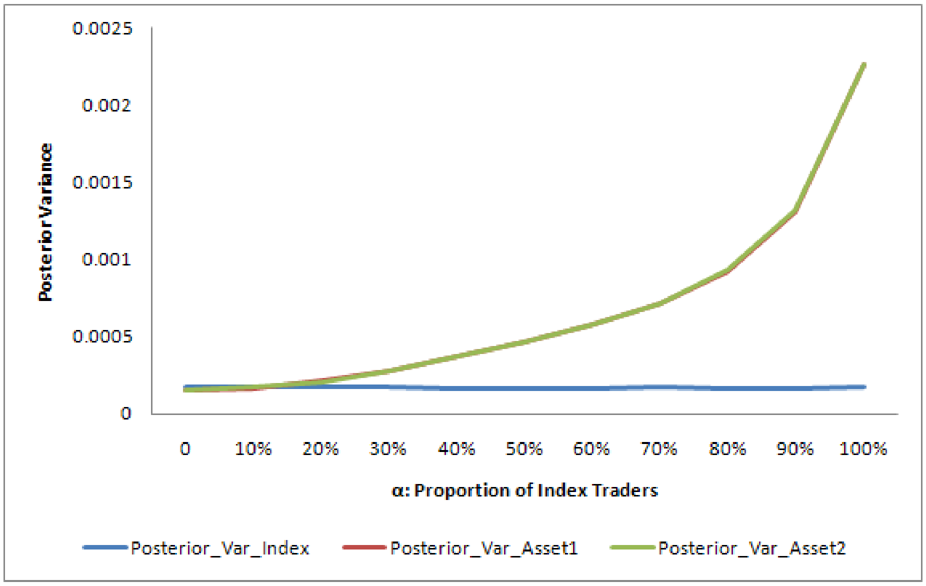


Figure 1.1: Relation between Posterior Variance of Return and Proportion of Index Traders

Table 1.1: Variables Related to Calculating the Posterior Probability

Variable	Description
p_{xj}	prior probability of risky asset j; $x = \{v_h, v_l\}$ and $j = 1, 2$.
p_{XI}	prior probability of index fund I; $V = \{V_H, V_M, V_L\}$.
q_{xj}	posterior probability of risky asset j; $x = \{v_h, v_l\}$ and $j = 1, 2$.
q_{XI}	posterior probability of index fund I; $V = \{V_H, V_M, V_L\}$.
π_{yxj}	risky asset j's conditional probability of y given x; $x = \{v_h, v_l\}$, $y = \{v_h, v_l\}$, and $j = 1, 2$.
π_{YXI}	index fund I's conditional probability of Y given X; $X = \{V_H, V_M, V_L\}$ and $Y = \{V_H, V_M, V_L\}$.
N_{Hs}	number of occurrence of a view on index equalling V_H in the set $O^{s\prime}$
N_{Ms}	number of occurrence of a view on index equalling V_M in the set $O^{s\prime}$
N_{Ls}	number of occurrence of a view on index equalling V_L in the set $O^{s\prime}$
N_{Hi}	number of occurrence of a view on index equalling V_H in the set O^i
N_{Mi}	number of occurrence of a view on index equalling V_M in the set O^i
N_{Li}	number of occurrence of a view on index equalling V_L in the set O^i
N_{hj}	number of occurrence of a view on asset j equalling v_h in the set O^s; $j = 1, 2$.
N_{lj}	number of occurrence of a view on asset j equalling v_l in the set O^s; $j = 1, 2$.

Table 1.2: S&P 500 Additions and Deletions

Date	Addition	Deletion	Reason
10/21/2013	Transocean	DELL Inc.	Dell acquired by private equity consortium
11/13/2013	Michael Kors	NYSE Euronext	ICE Exchange acquired NYSE Euronext
12/2/2013	Allegion	J.C. Penney	Allegion spun off by Ingersoll Rand
12/10/2013	General Growth Properties Inc.	Molex Inc.	MOLX acquired by Koch Industries
12/21/2013	Alliance Data Systems	Abercrombie & Fitch	Market capitalization changes
12/21/2013	Mohawk Industries	JDS Uniphase	Market capitalization changes
12/21/2013	Facebook	Teradyne	Market capitalization changes

Data Source: http://en.wikipedia.org/wiki/List_of_S%26P_500_companies

Table 1.3: Summary Statistics: S&P 500 Component Stocks in the Whole Sample

Variable	Mean	Std	10%	25%	50%	75%	90%
Realized_Vol_1min	0.0115	0.0077	0.0070	0.0083	0.0101	0.0127	0.0165
Realized_Vol_5min	0.0112	0.0067	0.0066	0.0079	0.0098	0.0126	0.0165
Index_Trading_Portion	0.2426	0.1586	0.0797	0.1299	0.2074	0.3157	0.4483
Trading_Dvol ($millions)	157.16	249.45	27.63	48.06	90.23	172.18	336.16
MktCap ($millions)	32.42	50.35	5.84	8.96	15.78	32.25	71.68

Table 1.4: Summary Statistics: S&P 500 Component Stocks within the Size Groups

Variable	Mean	Std	10%	25%	50%	75%	90%
Large Group							
Realized_Vol_1min	0.0104	0.0061	0.0065	0.0077	0.0093	0.0115	0.0145
Realized_Vol_5min	0.0100	0.0054	0.0060	0.0073	0.0090	0.0113	0.0145
Index_Trading_Portion	0.3013	0.1580	0.1273	0.1870	0.2736	0.3851	0.5089
Trading_Dvol ($millions)	300.54	369.75	81.82	122.95	198.08	347.31	593.88
MktCap ($millions)	73.37	70.50	25.99	32.05	47.85	82.54	157.33
Medium Group							
Realized_Vol_1min	0.0115	0.0082	0.0073	0.0085	0.0102	0.0126	0.0159
Realized_Vol_5min	0.0112	0.0067	0.0068	0.0081	0.0098	0.0124	0.016
Index_Trading_Portion	0.2304	0.1512	0.0831	0.1274	0.1965	0.2892	0.4131
Trading_Dvol ($millions)	107.40	114.38	36.63	54.14	82.55	124.72	189.29
MktCap ($millions)	16.19	3.63	11.85	13.36	15.71	18.59	21.39
Small Group							
Realized_Vol_1min	0.0127	0.0085	0.0075	0.0089	0.0110	0.0140	0.0188
Realized_Vol_5min	0.0123	0.0075	0.0070	0.0085	0.0107	0.0140	0.0187
Index_Trading_Portion	0.1951	0.1475	0.0565	0.0966	0.1588	0.2479	0.3726
Trading_Dvol ($millions)	61.63	60.95	17.93	27.30	44.85	74.92	120.97
MktCap ($millions)	7.19	2.22	4.17	5.36	7.30	8.92	10.08

Table 1.5: Summary Statistics: S&P 500 Index ETFs

Variable	Mean	Std	10%	25%	50%	75%	90%
Realized_Vol_1min	0.0047	0.0015	0.003	0.0036	0.0044	0.0056	0.0065
Realized_Vol_5min	0.0045	0.0017	0.0026	0.0032	0.0043	0.0055	0.0065
Index_Trading_Portion	0.2167	0.0519	0.1609	0.1741	0.2126	0.2539	0.2837
Index_Dvol ($millions)	17,136.85	5,639.65	10,579.17	13,366.82	15,450.45	20,624.04	24,868.99

Table 1.6: Price Discovery from S&P 500 Index to Component Stocks

Price_Type	Mean	Median
1-Minute_Midquote	25.69%	24.61%

Table 1.7: Price Discovery from Component Stocks to S&P 500 Index

Price_Type	Mean	Median
1-Minute_Midquote	21.57%	20.31%

Table 1.8: Regression Results for Realized Volatility of Stocks in the Whole Sample

Variable	Realized_Vol_1min	Realized_Vol_5min
Index_Trading_Portion	0.0088**	0.0063**
	(2.34)	(2.03)
log(Trading_Dvol)	0.0036***	0.0044***
	(5.2)	(4.86)
Adjusted R2	0.35	0.32

The coefficients, t-statistics, and R^2s are averages of the corresponding numbers of total sample stocks. ***, **, and * represent significance at the 1%, 5%, and 10% levels, respectively.

Table 1.9: Regression Results for Realized Volatility of Stocks within the Size Groups

Variable	Realized_Vol_1min	Realized_Vol_5min
Large Group		
Index_Trading_Portion	0.0072***	0.0075**
	(2.77)	(2.34)
log(Trading_Dvol)	0.0039***	0.0046***
	(6.21)	(5.77)
Adjusted R2	0.41	0.37
Medium Group		
Index_Trading_Portion	0.0102**	0.0102*
	(2.19)	(1.88)
log(Trading_Dvol)	0.0031***	0.0038***
	(4.90)	(4.55)
Adjusted R2	0.34	0.31
Small Group		
Index_Trading_Portion	0.0089**	0.0012*
	(2.05)	(1.85)
log(Trading_Dvol)	0.0038***	0.0047***
	(4.48)	(4.25)
Adjusted R2	0.31	0.29

The coefficients, t-statistics, and R^2s are averages of the corresponding numbers of stocks within each size group. ***, **, and * represent significance at the 1%, 5%, and 10% levels, respectively.

Table 1.10: Regression Results for Realized Volatility of the Index

Variable	Realized_Vol_1min	Realized_Vol_5min
Index_Trading_Portion	0.0007	0.0016
	(0.24)	(0.52)
log(Index_Dvol)	0.0039***	0.0043***
	(7.75)	(8.19)
Adjusted R2	0.66	0.7

***, **, and * represent significance at the 1%, 5%, and 10% levels, respectively.

CHAPTER 2

NEWS, INFLUENCE, AND EVOLUTION OF PRICES IN FINANCIAL MARKETS

2.1 Introduction

As social beings, we are influenced by many elements in our environment. This influence can cause us to behave in a manner that is suboptimal. The literature on information cascades (e.g., Banerjee (1992), Bikhchandani, Hirshleifer, and Welch (1992), and Welch (1992)) demonstrates how herding behavior can occur when agents follow others even when their private information suggests they shouldn't. A major source of influence in the life of investors is the media, or more specifically, opinions by experts and media pundits that are widely disseminated in various media forms that include both traditional outlets such as newspapers and television as well as newer forms such as Internet blogs and on-line forums. Such influence has the potential to help prices better adjust to information, but also could introduce shared errors into prices and detract from their informational efficiency.

Treynor (1987) claims that shared errors in published research are particularly important to asset prices. While the published opinion could be more accurate than the average accuracy of individuals' opinions, it replaces many independent estimates with a single number. Errors in the published opinion, therefore, will be reflected in the estimates of all investors it influences. Treynor suggests that the impact of such published opinions on asset prices would depend on their accuracy as well as the extent of their influence, or how persuasive they are.

Our objective in this paper is to study the influence of published opinions on the evolution of prices. Our model starts with an asset-specific news event. After the event, rational investors form personal "views" about the true value of the asset. Investors do not have full confidence in their perceptions of the value of the asset, and therefore they use Bayes' rule to combine their views with the market view. This is similar in spirit to Black and Litterman's (1992) definition of views as "feelings that some assets or currencies are overvalued or undervalued at current market prices." If errors in views are not systematically biased, then, at least formally, there is no difference between information and views. The interpretation of views, however, may be somewhat different. In a model with private information, the number of informed investors is often assumed to be small (sometimes just one), and information is considered accurate. Views, on the other hand, can be held by anyone, although the accuracy of personal views is considered to be marginal. Views can also give rise to the "wisdom of the crowds" effect whereby the aggregate behavior of many people with limited (but independent) information can bring about an accurate outcome (see, e.g., Surowiecki (2004)).

We assume that the analysis provided by experts and media pundits—the published view—is valuable in that it is more accurate than the average personal view. Still, the published view can also be incorrect, which is how shared errors are introduced into the price. In our sequential trade model, some investors are exposed to and influenced by the published view. This influence creates patterns in the trading process and the evolution of prices. The influence of the published view increases the price impact of trades (i.e., the trading costs) during an initial price-adjustment period. The reason for this result is that investors who seek (and are influenced by) the published view incorporate it into their trading, effectively imposing adverse selection on other investors. These increased costs, which last for a (random) number of trading rounds, deter regular investors who are not influenced by the published opinion, and they refrain from trading. In other words, their valuation is inside the spread when they arrive in the market, and hence they optimally choose not to trade.

As prices adjust to the published view, uncertainty decreases and the spread becomes smaller. From a certain point on, regular investors start trading on their views, and trading volume increases because both influenced and noninfluenced investors trade in the subsequent price-adjustment period, while only influenced investors trade in the initial price-adjustment period. This leads to changes in the evolution of prices during the subsequent price-adjustment period. In particular, the wisdom of the crowds effect emerges as the independent personal views of all investors get impounded into prices via the trading process. It is this wisdom of the crowds effect that ensures that prices converge to the true value, and also that shared error introduced by the published view is ultimately expunged from prices.

This wisdom of the crowds effect, which ultimately corrects prices, sets the model apart from the information cascade papers. In particular, once investors in information cascades models start relying on common information rather than their own signals, they continue to do it indefinitely unless some exogenous influence breaks that dependency. This exogenous influence can be the release of public information or the introduction of a class of better-informed investors. In our model, the endogenous reduction in trading costs increases the pool of regular investors who trade the asset after some time passes, and the confluence of their marginally informative views delivers the wisdom of the crowds effect and corrects prices.

Of course, along the way, the influence of published opinions changes the evolution of prices and creates volume patterns in the market. The more influence is exerted by the published view, the more impact it has on the trading process and the resulting prices. Treynor (1987) states that the number of investors influenced by the published opinion increases with the time elapsed since publication. We use this insight to empirically test the implications of our model. Our empirical tests use corporate earnings announcements as the information events. These are often associated with heightened media attention in traditional media outlets as well as in blogs, on-line discussion groups, and investor newsletters. Various experts and media pundits use this opportunity to voice their opinions on firms, which reflect their interpretations of the information in the earnings announcements. The passing of time following an earnings announcement provides opinion writers with more opportunities to distribute their views, and increases the likelihood that investors get exposed to and are influenced by these views.

Almost all earnings announcements take place either in the morning, before the market opens, or in the afternoon, after the main trading session on organized exchanges closes. Morning announcements are followed very quickly by the opening of the market, decreasing the likelihood that published opinions are produced and many investors are exposed to them before they submit their orders to trade at the opening of the regular trading session. For afternoon announcements, there is much more time for published opinions to be generated and influence investors before the investors submit orders for the regular trading session on the following day.[1] We therefore compare how prices evolve for morning versus afternoon earnings announcements to test the implications of our model on how the influence of published opinions affects the trading process.

[1]While investors could potentially trade after-hours (i.e., outside the 9:30 a.m.–4:00 p.m. regular trading session), the illiquid nature of trading outside the regular trading session deters many investors. Most investors, therefore, have a strong preference for trading during the regular trading session, providing published views on afternoon announcements with more opportunity to reach and influence investors.

It is reasonable to assume that any published opinion following the earnings announcement would reflect or rely on the earnings surprise (i.e., how far the earnings number is from the consensus analysts' forecast prior to the announcement). Therefore, we use the earnings surprise as a proxy for the nature of the published view. To test the implications of our model, we match pairs of morning (less influence) and afternoon (more influence) announcements on the strength of the published view as well as various attributes of the stocks.

The first implication we test is that the initial price adjustment (or price impact) is larger when there is stronger influence of the published view on investors. We look at both the absolute value of close($t-1$)-to-open(t) returns and at Amihud's price impact measure, and find evidence consistent with our model. The most important implications of the model, however, involve the dynamic adjustment of prices. In the presence of influence, positive published views would generate a large positive return during the initial price-adjustment period. As the wisdom of the crowds effect emerges, however, the return over the subsequent price-adjustment period could exhibit one of two patterns: (i) a reversal, if the published view was incorrect and prices adjust downward, or (ii) a small positive return as the independent views of additional investors complete the adjustment of prices upwards. In either case, the difference in return between the subsequent and initial price-adjustment periods should be negative. A similar logic suggests that for negative published views, this subsequent-minus-initial return difference in return should be positive. We find that indeed these return patterns are stronger in afternoon announcements than in morning announcements, in line with the predictions of the model.

In addition to price patterns implications, the influence model also yields a volume implication. In particular, some investors (who are not influenced by the published view) optimally choose to refrain from trading during the initial price-adjustment period due to the higher trading costs. Therefore, there is lower volume initially as prices adjust to the published opinion, and volume increases when the wisdom of the crowds effect emerges and all investors join the trading process during the subsequent price-adjustment period. We test this prediction, comparing afternoon to morning earnings announcements, and find results consistent with the theory. Hence, our results suggest that in addition to the evolution of prices, published views also affect the trading process itself.

Besides the literature on influence and information cascades that we mentioned at the beginning, our study is also related to the literature on the timing of the release of earnings information (for example, during the trading day versus after the market closes, or during the week versus on Friday). Patell and Wolfson (1982) find that bad news is more often released after the market closes and suggest a couple of possible explanations. First, managers could opportunistically release bad news after the market closes at a time of reduced media coverage and investor attention. Second, managers could release news after the market closes to allow a longer period for dissemination and evaluation of the news.[2] Doyle and Magilke (2009) find evidence consistent with the latter explanation, and they also provide evidence that more complex firms tend to release earnings after the market closes.

Jiang, Likitapiwat, and McInish (2012) examine price discovery following earnings announcements in after-hours trading and during the regular trading session. They observe that price adjustment to announcements made after the close is greater than to announcements made in the morning before the market opens, which is consistent with our first finding. They do not observe a difference in the efficiency of price discovery between announcements made in the morning and in the afternoon. Michaely, Rubin, and Vedrashko (2014) find that the initial price impact (an hour after an announcement) is smaller for announcements made during the trading day compared with outside regular trading hours. They suggest that not all investors follow the market continuously, and hence announcing earnings outside the regular trading hours results in better price discovery because investors are given more time to evaluate the news.[3] This is consistent with our finding that the initial price impact of afternoon announcements is greater than that of morning announcements. Our study suggests that this effect can arise due to the influence of published opinions on investors.

The remainder of the paper is structured as follows. Section 2.2 presents the theoretical model. We describe the economy, characterize the equilibrium, and provide propositions on price impact and the dynamic evolution of prices and volume that describe how influence affects the trading process. These propositions also provide the basis for the empirical tests

---

[2]Gennotte and Trueman (1996) model the decision of a manager over the timing of his firm's earnings release. In their model, the price response is greater if the announcement is made during trading hours rather than after the market closes. The driving force behind this result is that there is greater likelihood of trades coming from informed traders during trading hours, while postponing trading to the following day increases the likelihood of additional noise trading.

[3]Michaely, Rubin, and Vedrashko (2014) also investigate the relation between corporate governance and the timing of earnings releases. They find that firms with poor corporate governance are more likely to release the news during the trading day.

of the model. Section 2.3 puts forward the empirical methodology, describing the sample and the matching procedure, while Section 2.4 presents the empirical results of the tests. Section 2.5 concludes the paper.

2.2 The Model

2.2.1 The Economy

We study a variant of the Glosten and Milgrom (1985) sequential trade model for a single asset with the interest rate set to zero. Before trade commences, the consensus is that the true value of the asset, denoted by $\tilde{v}$, is equally likely to be zero or one. Competitive and risk-neutral market makers set the ask and bid prices. Traders show up sequentially, and each trader has a personal view about whether the true value is zero or one. We call a view that the asset is worth one (zero) a positive (negative) view. The view of the n-th trader (i.e., the trader who arrives in period n) can be written as:

$$\tilde{v}^n = \tilde{v}(1 - \tilde{\epsilon}^n) + (1 - \tilde{v})\tilde{\epsilon}^n,$$

where the error term, $\tilde{\epsilon}^n$, takes the value zero if the view is correct and one otherwise. The probability that a personal view is correct, denoted by π, is identical for all traders and satisfies $\pi > 1/2$.

There are two classes of traders: knowledgeable investors and noise traders. The probability that a knowledgeable investor shows up at any given period is μ, and the probability that a noise trader arrives in the market is $1 - \mu$. Noise traders, irrespective of their views and for reasons that are exogenous to the model (e.g., risk sharing, liquidity needs), either buy or sell with equal probabilities. Knowledgeable investors, on the other hand, consider their personal views when making a decision whether to buy, sell, or abstain from trading. A fraction $\alpha \geq 0$ of the knowledgeable investors seek out the opinions of experts and media pundits. We call the consensus opinion propagated by these experts the "published" opinion. The fraction α, therefore, represents the extent of influence of this published opinion. The published opinion is also associated with a view about the asset's true value that we denote by $\tilde{v}^e$ and let $\tilde{\epsilon}^e$ denote the error associated with the view, i.e.,

$$\tilde{v}^e = \tilde{v}(1 - \tilde{\epsilon}^e) + (1 - \tilde{v})\tilde{\epsilon}^e.$$

The probability that the published expert view is correct, denoted by π^e, is strictly greater than π, the probability that an investor's personal view is correct. However, the published view can also be incorrect (with probability $1 - \pi^e$), and this is important to the intuition discussed by Treynor (1987) regarding how such experts and media pundits can introduce errors into the price.

We call knowledgeable investors "influenced" if they seek and pay attention to the published opinion and "regular" if they do not. We assume that errors in views, the true value of the asset, the probabilistic selection model that governs the choice of traders, and the trading decision of noise traders are all independent.

As in traditional sequential trade models (e.g., Glosten and Milgrom (1985) and Easley and O'hara (1992)), each trading period is long enough to accommodate at most one trade. When a trader arrives in period n, he has the option of buying one unit of the stock at the quoted ask price, selling one unit at the quoted bid price, or abstaining from making a trade. We let H_n denote the history of trading up to the n-th period. Given the ask and bid prices, the expected profit of a regular investor is:

$$\begin{cases} E[\tilde{v}|H_n, \tilde{v}^n] - ask_n & \text{if buys} \\ 0 & \text{if abstains} \\ bid_n - E[\tilde{v}|H_n, \tilde{v}^n] & \text{if sells} \end{cases} \tag{2.1}$$

and the expected profit of an influenced investor is:

$$\begin{cases} E[\tilde{v}|H_n, \tilde{v}^n, \tilde{v}^e] - ask_n & \text{if buys} \\ 0 & \text{if abstains} \\ bid_n - E[\tilde{v}|H_n, \tilde{v}^n, \tilde{v}^e] & \text{if sells.} \end{cases} \tag{2.2}$$

An *equilibrium* is a bid_n and ask_n quote (a mapping from H_n to R), such that (i) given the quoted prices, investors maximize their profits, and (ii) market makers quote the following regret-free prices:

$$ask_n = E[\tilde{v}|H_n, buy]$$

$$bid_n = E[\tilde{v}|H_n, sell].$$

Given the realizations of all random variables in the model, we are interested in contrasting the equilibrium outcomes in two environments: $\alpha > 0$ and $\alpha = 0$. We refer to the former as the "influence" model, wherein a portion of investors is influenced by experts and media pundits, while the latter is the "benchmark" model. In the benchmark model, all knowledgeable investors are regular investors.

2.2.2 Properties of the Equilibrium

We can classify traders into eight types based on the source and realization of their views. Table 2.1 provides a taxonomy of the types, which we denote by $\theta = 1, ..., 8$. For example, type $\theta = 2$ is an influenced investor whose personal view is positive but is exposed to a negative published view. Types 3 and 4 represent regular investors who pay attention to their personal views only, while types 7 and 8 are the noise traders who buy and sell for exogenous reasons. We let $\tilde{\theta}_n$ denote the (random) type that arrives to the market in period n. Knowing which type is picked is informationally equivalent to the information that the trader knows. For example, knowing that $\tilde{\theta}_n = 1$ is equivalent to knowing that both the investor's own view as well as the published view are positive ($\tilde{v}^n = 1$ and $\tilde{v}^e = 1$).

Proposition 2 *There exists a Markovian equilibrium with three state variables.*[4]

$$\begin{aligned} p_n^h &\equiv P(\tilde{v} = 1 | H_n, \tilde{v}^e = 1) \\ p_n^m &\equiv P(\tilde{v} = 1 | H_n) \\ p_n^l &\equiv P(\tilde{v} = 1 | H_n, \tilde{v}^e = 0) \end{aligned} \tag{2.4}$$

For every finite n, and regardless of how the history of trading unfolds, the state variables remain in the open interval $(0,1)$ and the bid-ask spread remains strictly positive. In the benchmark model (i.e., $\alpha = 0$), p_n^h and p_n^l are redundant.

Proposition 3 *As n goes to infinity, p_n^m converges to $\tilde{v}$ almost surely.*

Proofs of all the propositions are provided in Appendix B. In the Markovian equilibrium, regular investors interpret the history of trading in the exact same manner as market makers do, and then weigh in with their personal views. In contrast, influenced investors interpret the trading history differently. This is because influenced investors know that some types of traders are not present in the market. For example, if $\tilde{v}^e = 1$ (i.e., the published view is positive), then influenced investors know that types five and six are not present in the market. Table 2.2 shows the valuation each trader type attaches to the asset, and how they can be expressed as functions of the three state variables. This choice of the state variables enables us to write the valuations, which are equivalent to conditional probabilities because the asset value can only take the values zero or one, in an especially simple form in the Markovian equilibrium.

---

[4]The initial conditions of the state variables are given by

$$\begin{aligned} p_0^h &= \pi^e \\ p_0^m &= 1/2 \\ p_0^l &= 1 - \pi^e. \end{aligned} \tag{2.3}$$

Our goal is to use the model to investigate the patterns in prices and trading created by the influence of experts and media pundits on investors. For that purpose, the next two propositions contrast the influence model and the benchmark model.

Proposition 4 *When trading opens (i.e., at $n = 1$), the initial price impact of orders in the influence model is larger than the initial price impact in the benchmark model.*

In a sequential trade model, the price impact is equivalent to the updating of beliefs about the asset value brought about by the order flow.[5] Due to the greater precision of the beliefs that incorporate the published view, the initial price impact is larger in the influence model. This, irrespective of whether the published view is correct or not, is the reason an incorrect published view impacts prices more than an incorrect personal view. The most interesting insights of the model come from the dynamic nature of trading and price adjustment that are investigated in the next proposition. Let v_θ be the valuation of a trader of type θ (from Table 2.2).

Proposition 5 *Assume the probability that a personal view is correct is sufficiently small (i.e., π is close to 1/2). In the influence model (i.e., when $\alpha > 0$), there exists an $\tilde{N} \geq 1$ such that for all $n \leq \tilde{N}$*

1. *p_n^h and p_n^l remain constant, and $p_n^m = E[\tilde{v}^e | H_n]$.*
2. *The bid-ask spread is larger than $v_3 - v_4$ (i.e., the ask (bid) is higher (lower) than the valuation of a regular investor with a positive (negative) view).*
3. *Regular investors abstain from trading, and influenced investors trade in the direction of the published view regardless of their personal views.*

For $n > \tilde{N}$, we can show that for sufficiently large n,

D) *The bid-ask spread is smaller than $v_3 - v_4$ (i.e., the ask (bid) is lower (higher) than the valuation of a regular investor with a positive (negative) view).*

E) *All investors trade when they arrive in the market.*

In the benchmark model ($\alpha = 0$), arriving knowledgeable investors never abstain from trading and always trade in the direction of their personal views. In particular, the bid-ask spread is always smaller than $v_3 - v_4$.

---

[5] The bid-ask spread in the model is simply the sum of the price impacts for buying and selling.

Trading costs—which in the sequential trade model are simply the price impacts—are higher when the market opens and in the early trading periods. Proposition 5 demonstrates the difference influence makes for the evolution of prices and volume in the market. These trading costs during that initial price-adjustment period (when $n \leq \tilde{N}$) deter regular investors from participating and they refrain from trading (i.e., their valuation is inside the spread when they arrive in the market and hence they optimally choose to refrain from trading). As prices adjust to the published view, uncertainty decreases and the spread narrows. After some time, regular investors start to trade on their views, and this leads to changes in price and volume patterns. First, volume increases due to the fact that all investors trade (previously only the influenced investors traded). Second, the wisdom of the crowds effect emerges as the independent views of all investors get impounded into the price, and ensures convergence to the true value (in Proposition 3). The convergence may still be slowed by the published view's impact on the valuations of the influenced investors, but the wisdom of the crowds effect eventually prevails.

To provide a feel for the result, Figure 2.1 shows the evolution of prices when the published view is incorrect: it is negative while the true value of the asset is one. Panel A focuses on the first 500 periods and Panel B continues the simulation up to period 3,000. The red line in the figure shows the simulated prices in the influence model while the dotted blue line is from a simulation of the benchmark model. The parameters we use for both simulations are $\pi = 0.55$, $\pi^e = 0.9$, and $\mu = 0.5$. We take $\alpha = 0.5$ for the influence model, while $\alpha = 0$ by definition in the benchmark model. Panel A clearly shows the larger initial price impact of order flow in the influence model compared with the benchmark model. Prices adjust downward rapidly towards the valuation implied by the published view. They then move around that level for a while, until the independent personal views of investors become more dominant and the wisdom of the crowds effect brings prices up. Given that the published opinion is still reflected in the valuations of the influenced investors, however, convergence to the true value appears slower in the influence model, and only in panel B do we observe that prices converge to the true value.

Our maintained assumption is that the analysis provided by experts and media pundits—the published view—is valuable in that it is more accurate than the average personal view. It is this difference in accuracy that makes prices adjust very rapidly at the beginning of trading in the influence model. Still, the figure demonstrates that the published view can slow down the eventual convergence of prices to full information values. When the published view is correct, there is also a larger price impact in the influence model relative to the benchmark model at the beginning, and it is followed by a somewhat muted price adjustment later on. Irrespective of whether a negative published view is correct or not, the initial return in the influence model (i.e., the return from the prior expected value to the price at the market open or at any period $n \leq \tilde{N}$) is very negative while the return over the subsequent price-adjustment period (i.e., from any $n \leq \tilde{N}$ to the convergence of prices to the true value) is either positive as in Figure 2.1 or negative but of smaller magnitude.

Figure 2.2 shows the ratio of trading volume in the influence model to trading volume in the benchmark model in the first 100 periods of the same simulation. Specifically, we aggregate volume in 10-period buckets $(1-10, 11-20, ..., 91-100)$, and present the influence-to-benchmark ratio of trading volume in each bucket. In the benchmark model, all traders trade when they arrive in the market, and hence volume in each bucket is equal to ten. Consistent with Proposition 5, volume in the influence model is lower than in the benchmark model during the initial price-adjustment period $(n \leq \tilde{N})$ when regular investors abstain from trading, but it picks up after about 70 periods, from which point on volume is the same in the influence and benchmark models.[6]

One of the interesting insights that our theory yields, therefore, is the dynamic way in which influence of published views impacts market prices. In the model, trading by influenced investors can introduce errors into the price. Regular investors, who at the beginning (optimally) choose to wait on the sideline, start trading after the information in the published view is impounded into the price. The trading on independent personal signals, which we call the wisdom of the crowds effect, ultimately drives prices to the true value. This natural correction mechanism differentiates the implications of our model from those of models in the information cascades literature (see, for example, Banerjee (1992), Bikhchandani, Hirshleifer, and Welch (1992), and Welch (1992)), where agents' reliance on common information (rather than their own signals) continues indefinitely unless some exogenous influence is introduced (e.g., a public news announcement or a new agent type with more precise information). In our model, in contrast, the influence of the published opinion eventually disappears and we observe the wisdom of the crowds emerging to impact the price path endogenously without the need for an exogenous intervention.

[6] How long regular investors abstain from trading depends on the parameters of the model and the specific realization of the sequence of trader types who arrive in the market.

2.3 Empirical Methodology

Our model enables us to examine the influence of expert published opinions on financial markets. The more persuasive these published opinions are, the greater the impact they have on the evolution of prices. Our empirical tests use corporate earnings announcements as identifiable news events. These are often associated with much discussion in media outlets as well as Internet bulletin boards and investor newsletters. Various pundits and financial newsletter writers use this information to update their audiences on their views with respect to the firms that announce the earnings numbers. Treynor (1987) notes that the number of investors persuaded by a piece of published analysis or opinion increases with the time elapsed since publication. This insight suggests that time elapsed following the earnings announcement gives more opportunity to opinion writers to distribute their views, and to investors to be exposed to these views and become persuaded.

Earnings announcements tend to take place either in the morning, before the stock market opens, or in the afternoon, after the regular trading session on the organized exchange closes. For morning announcements, it is less likely that published views are produced and many investors are influenced by these views before the regular trading session opens. We expect that the orders of investors in this case more likely reflect their own personal views. These morning announcements are therefore closer to our theoretical benchmark model (where $\alpha = 0$). In all likelihood, α is positive for morning announcements but rather small, as published opinions could be issued before trading begins but reach only a small number of investors. Afternoon announcements are different in that there is an evening, night, and morning before the main trading session on organized exchanges opens, providing more time for published opinions to be generated and influence investors. While there is "after-hours" trading in the U.S. in which investors can potentially trade on their views before the opening of the main trading session on the following day, volume in after-hours trading is low in general and markets are illiquid.[7] Many investors, therefore, have a strong preference for trading during the regular trading session, providing more time for their views to be influenced by published opinions. We take these afternoon announcements to represent the influence model of our theory (where $\alpha > 0$). In other words, we use the insight from Treynor (1987) that the number of investors influenced by a published opinion increases with time as the basis for our maintained assumption that

[7] Jiang, Likitapiwat, and McInish (2012), however, note that more after-hours trading occurs when an earnings announcement takes place than on nonannouncement days.

$\alpha_{\text{Afternoon}} > \alpha_{\text{Morning}}$.[8] We therefore compare how prices evolve for afternoon versus morning earnings announcements to test the implications of our theory on the difference between the influence and benchmark models.

We use the sample of corporate earnings announcements from Doyle and Magilke (2009). The sample period is 2000 through 2005, and the sample identifies both the date of the earnings announcement as well as whether the announcement occurs in the morning (before the market opens at 9:30 a.m.) or in the afternoon (after the market closes at 4:00 p.m.). There are 26,443 morning announcements and 23,893 afternoon announcements in the dataset. We merge the earnings announcement dataset with three additional data sources: CRSP, I/B/E/S, and TAQ. The merging of datasets results in the loss of some observations, and we are left with 25,008 morning and 21,848 afternoon announcements.

We need a proxy for the published view, or the news content, of each announcement. It is reasonable to assume that any published opinion of pundits or newsletter writers at the time the announcement is made would rely on the earnings surprise (i.e., how far the earnings number is from the consensus analysts' forecast prior to the announcement). We therefore use a standardized earnings surprise measure (ES), defined as the difference between actual earnings and the mean consensus analysts' forecast from I/B/E/S divided by the price of the stock a month prior to the announcement, to sort the earnings announcements into seven categories.[9]

All positive ES announcements are sorted into three equal-sized categories, whereby category 1 contains the strongest positive earnings surprises and category 3 the weakest positive earnings surprises. Category 4 contains all earnings announcements with zero ES. All negative ES announcements are sorted into three equal-sized categories: category 5 contains the weakest negative earnings surprises and category 7 the strongest, most negative, earnings surprises. In the model we assume that the published view is either very good ($v^e = 1$) or very bad ($v^e = 0$). We therefore carry out the empirical work on categories

[8] Engelberg and Parsons (2011) show that local newspaper coverage of an earnings announcement increases trading from local retail investors in their sample of clients of a discount broker. Of course, only afternoon announcements from the previous day can make it into the newspaper, suggesting that the influence of published opinions would in fact be greater for afternoon announcements than for morning announcements.

[9] It is likely that α is positive to some extent in all earnings announcements, and hence we use ES as a proxy for the published view for both afternoon and morning announcements. We chose the terminology we use in the paper—comparing the influence versus benchmark models—to simplify the exposition. All the implications of the model that we discuss carry through when the influence model has two levels of α such that $\alpha_{\text{Afternoon}} > \alpha_{\text{Morning}}$.

1 and 2, representing the positive published view in the model, and categories 6 and 7, representing the negative published view in the model.[10] Table 2.3 provides information on the number of morning and afternoon earnings announcements in the entire sample as well as separately in ES categories 1, 2, 6, and 7.

To properly compare the predictions of the benchmark and influence models, we need to match pairs of morning and afternoon announcements to neutralize differences in the strength of the published views as well as in other attributes of the stocks. Since we use ES to represent the published opinion, we control for the strength of the view by matching morning to afternoon announcements only within the same ES category (1, 2, 6, or 7). The first general attribute of stocks that we match on is industry, and we classify each earnings announcement into 10 industries using the classification method developed by Kenneth French.[11] This second step leaves us with 40 ES/Industry cells. Table 2.4 provides the number of earnings announcement pairs in each ES/Industry cell as well as in the entire sample. Outside of the "Other" classification, the largest number of pairs can be found in the Business Equipment classification that includes computers, software and electronic equipment. Consumer Durables as well as Telephone and Television Transmission are two of the smaller industries.

Davies and Kim (2009) discuss the merits of various matching procedures, and their analysis suggests matching by market capitalization and price as the attributes of stocks. Therefore, within each ES/Industry cell, we match a morning earnings announcement with an afternoon earnings announcement (with replacement) by choosing the afternoon announcement that minimizes the distance function:

$$\left(\frac{Price_i - Price_j}{Price_i + Price_j}\right)^2 + \left(\frac{MktCap_i - MktCap_j}{MktCap_i + MktCap_j}\right)^2 \tag{2.5}$$

where i denotes the morning announcement and j denotes an afternoon announcement. Table 2.5—Table 2.8 provide percentile summary statistics for the matched pairs. Table 2.5 and Table 2.7 present market capitalization, price, and ES of the morning and afternoon earnings announcements in the matched pairs of ES categories 1 and 2 (6 and 7). We observe a rather close matching on these three attributes. To see whether moving from four ES

[10]The results of the tests in Section 2.4 are similar in nature when we use just the extreme categories (1 and 7).

[11]The mapping of four-digit SIC codes into the 10 industries can be found on Kenneth French's web site: http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data Library/det 10 ind port.html

categories to 40 ES/Industry cells affects the quality of matching on market capitalization and price, we provide in Table 2.6 and Table 2.8 summary statistics for an alternative matching scheme in which we proceed directly to match on market capitalization and price within each ES category without controlling for industry. The matching on market capitalization in the 90th percentile is slightly better when we do not match on industry first, but otherwise controlling for industry does not appear to impact the market capitalization and price matching. We therefore present in the paper the empirical analysis using the matching procedure that controls for ES, industry, market capitalization, and price.[12]

Throughout the empirical investigation, we test for differences between afternoon announcements (representing the influence model) and morning announcements (representing the benchmark model) using pairs' tests. We report the mean and median of the paired differences between the afternoon and morning announcements together with p-values from a pairs' t-test and a Wilcoxon signed-rank test against the two-sided hypothesis of zero differences.[13]

2.4 Empirical Results

2.4.1 Price Impact

The first implication that we test empirically (from Proposition 4) is that the initial price adjustment, or price impact, is larger in the influence model, where investors listen to and rely on the published opinion, than in the benchmark model. Time elapsed since the earnings announcement gives more opportunity for expert analysis and opinion to be distributed and for investors to be influenced, and so the initial price adjustment should be larger in afternoon announcements than in morning announcements. The intuition behind this prediction of the model is that whether or not the published opinion is correct, the price impact is driven by the precision of the published view, which is greater than that of the average personal view of investors.

For an earnings announcement on day t, we take the closing price on day $t-1$ as the

---

[12] We carried out the empirical analysis presented in Tables 2.9 through 2.15 on the alternative matching scheme that does not control for industry, and found that the results were similar to those we present in the paper. The results without the industry control are available from the authors.

[13] Only a negligible number of pairs (e.g., eight pairs in category 6) had both afternoon and morning announcements sharing the same date. Similarly, there were only a very small number of pairs in each category that consisted of the same two stocks. Therefore, we did not carry out clustering of the errors by date or firm.

price that reflects the prior beliefs before the earnings announcement. The closing price is defined as the midquote (ask plus bid divided by two) for the quote in the TAQ database with MODE=3 (Closing Quote) on day $t-1$ from the market on which the stock is listed, or if such a quote does not exist, the midquote prevailing at 4:00 p.m.[14] For the price that reflects the initial price impact after the information event and the dissemination of the published view we take the opening price on day t. The opening price is defined as the midquote with MODE=10 (Opening Quote) from the market on which the stock is listed, or if such a quote does not exist, the first quote after 9:30 a.m..

We test the magnitude of the initial price adjustment in two ways. Our first test looks at the differences between afternoon and morning announcements in the magnitude (or absolute value) of the close-to-open return, AbsRet. According to the model, greater influence by the published view implies a larger initial price adjustment and hence the difference in AbsRet between the influence and benchmark models should be positive for all ES categories. For our second test, we divide AbsRet by dollar volume to create Amihud's measure of price impact.[15] Due to after-hours trading and because we do not know the exact time of the afternoon earnings announcement, we need to make a choice as to the time from which we begin aggregating volume. We choose 6:00 p.m. based on evidence that most afternoon earnings announcements occur between the close of trading and 6:00 p.m. (see Jiang, Likitapiwat, and McInish (2012)), and hence the volume that accumulates from 6:00 p.m. on is likely to reflect order flow that arrives after the announcement. We include the volume at the open (the opening auction for NYSE stocks or the opening trade for NASDAQ stocks) for all earnings announcements. Since Amihud's Measure requires dividing by volume, we exclude announcements with very little volume in order to minimize the number of outliers.[16]

The first line of Table 2.9 presents the results for AbsRet. We observe that the difference

---

[14] Following Boehmer, Saar, and Yu (2005), we apply certain filters to the raw quote and trade data to minimize data errors. Specifically, we keep only quotes with TAQs MODE field equal to 0, 1, 2, 3, 6, 10, 12, 23, 24, 25, and 26. We eliminate quotes with nonpositive ask or bid prices, or where the bid price is higher than the ask price. We keep trades for which TAQs CORR field is equal to either zero or one, and for which the COND field is either blank or equal to B, J, K, S, or E. We also exclude trades with nonpositive prices.

[15] Following Amihud (2002), we multiply the measure by 10^6.

[16] Specifically, we exclude an earnings announcement if the volume is less than \$10,000. We used other methods for robustness (e.g., winsorizing the measure at 2%) and the results were not sensitive to the choice of method by which we exclude outliers.

between afternoon and morning announcements is positive and statistically different from zero both for positive announcements (categories 1 and 2) and negative announcements (categories 6 and 7). The second line of Table 2.9 presents the results for Amihud's Measure. Here as well, the results are positive and highly significant using both the t-test and the Wilcoxon signed-rank test, although the fact that the mean is much larger than the median suggests that the division by volume creates some outliers. Using both measures, therefore, we find that the initial price adjustment is larger for afternoon announcements, which is consistent with the idea that more time allows the published opinion to reach and influence more investors and hence increases the magnitude of the initial price impact.

2.4.2 Return Patterns

After the initial adjustment of prices to the published view, regular investors who were on the sideline start trading, and trading on independent personal views brings about eventual convergence to the true value. If a negative published view is incorrect, as in Figure 2.1, we see fast initial price adjustment downward in the influence model, and therefore the initial return is very negative. Once prices incorporate the published view and regular investors start trading, the wisdom of the crowds effect emerges and we observe a reversal as prices adjust upward until they reach one (in Panel B), and hence the subsequent return is positive. RetChg, defined as the return over the subsequent price-adjustment period minus the return over the initial price-adjustment period, is therefore very positive, indicating the reversal. In contrast, the benchmark model shows a slower initial adjustment and a continuation in the same general direction (upward) rather than a reversal. RetChg for the benchmark model would therefore be very small.[17] Hence, the difference between afternoon announcements (the influence model) and morning announcements (the benchmark model) when the published view is incorrect should be positive in ES categories 6 and 7 (which represent negative published views).

If the published opinion is correct, there is faster initial adjustment downward in the influence model than in the benchmark model. The subsequent price adjustment to the true value in the influence model is both small (because much of the price change occurred in the initial price-adjustment period) and slow (due to the tight posterior beliefs). Therefore,

[17] The exact magnitude of RetChg for the benchmark model depends on the length of time assumed for the initial price-adjustment period in the influence model. In general, it can be slightly positive, zero, or slightly negative.

RetChg(Afternoon) would be positive due to the strictly concave curvature of the price adjustment, while RetChg(Morning) would be close to zero. Hence, the difference between afternoon and morning announcements when the published view is correct should also be positive. Let DiffRetChg(6&7) be defined as:

$$\text{DiffRetChg}(6\&7) = \text{RetChg(Afternoon)} - \text{RetChg(Morning)} > 0. \tag{2.6}$$

Since DiffRetChg(6&7) would be positive in the influence model irrespective of whether the published view is correct or not, this prediction lends itself nicely to testing using our data. For positive published views (ES categories 1 and 2), one can follow the same logic to arrive at the following relationship:

$$\text{DiffRetChg}(1\&2) = \text{RetChg(Afternoon)} - \text{RetChg(Morning)} < 0. \tag{2.7}$$

To test these predictions, we need to define the return intervals for the initial and subsequent price-adjustment periods. We use several definitions to ensure that our findings are robust. Our main definition for the initial price adjustment is the return from previous day close to the opening of the market subsequent to the announcement (the close-to-open return, as in Section 2.4.1). A somewhat more arbitrary decision concerns the point at which prices have already adjusted to the information and converged to the true value. Therefore, we use several return intervals to ensure that our results are not sensitive to the particular choice of an endpoint for the subsequent price adjustment. Specifically, we compute four different returns: open-to-10:30 a.m., open-to-11:00 a.m., open-to-11:30 a.m., and open-to-12:00 p.m. using midquotes from the primary market at 10:30 a.m., 11:00 a.m., 11:30 a.m., and 12:00 p.m. to represent alternative points that are far enough into the subsequent price-adjustment period.

Table 2.10 shows the mean and median of DiffRetChg separately for positive and negative earnings surprises. For categories 6 and 7, DiffRetChg is positive and statistically significant as predicted in Equation (2.6), while DiffRetChg is negative and significant as in Equation (2.7) for categories 1 and 2. These results are therefore consistent with the predictions of the model pertaining to differences in the evolution of prices between the influence model (afternoon announcements) and the benchmark model (morning announcements). The magnitude of the effect appears twice as large for negative earnings announcements, perhaps suggesting a more dramatic initial reaction to negative published views.

The results in Table 2.10 reflect differences in raw returns between afternoon and morning announcements. However, the matched announcements in each pair do not generally happen on the same day, and therefore it could be that they reflect differences in market return alongside the effect predicted by our theory. To separate them, we repeat the analysis using excess returns computed by subtracting from each return measure the return on the largest ETF that follows the S&P 500 index (SPY) as a proxy for the market return.[18] The results of the tests using excess returns are provided in Table 2.11. A comparison of the two panels reveals that most of the effect remains intact, and the numbers have similar magnitudes and statistical significance.

Our third analysis seeks to ensure that our results are not driven by unknown firm attributes that may impact the return and that are not perfectly controlled for by our matching procedure. Specifically, for each earnings announcement we calculate the close-to-open or open-to-endpoint returns in the same stock in each of the four weeks prior to the announcement on the same day of the week as the earnings announcement. We average these four return observations to get the "normal" return of the stock, and subtract that normal return from the raw return to get an abnormal return measure for the earnings announcement. Table 2.12 provides the results for the abnormal return measure, and we observe a significant negative DiffRetChg for categories 1 and 2 and a significant positive DiffRetChg for categories 6 and 7. The results from Table 2.10 to Table 2.12 are consistent with the prediction of our theory, and they appear very robust to both various endpoints that we consider for the subsequent price-adjustment period as well as to multiple definitions of returns.

One interesting feature of the influence model that can be observed in the simulations in Figure 2.1 is that the initial price adjustment to the published view continues after the market opens for a short period of time. The reason for this pattern is that prices adjust to the published view through the trading of sequentially arriving investors, not instantaneously via some sort of mysterious coordination of beliefs. While the adjustment of prices on the first trade is large, it takes up to 70 periods in the simulation for prices to adjust to the published view before emphasis shifts to the independent personal views and prices begin to reflect the wisdom of the crowds. If the price in the first period in the figure is equivalent to the price at the opening of the regular trading session in the empirical analysis, DiffRetChg should be somewhat smaller than if we had used an alternative price

---

[18]SPY is heavily traded on multiple trading venues that are not its primary market. Therefore, we compute returns for SPY using the midquote of the National Best Bid and Offer (NBBO) that we construct using the quote information in TAQ.

after the opening of the market that reflects some additional trading. Our choice of the opening price for the tests in Table 2.10—Table 2.12 was meant to be conservative because it is difficult to ascertain when prices finish their initial adjustment to the published view and the focus shifts to the independent signals of investors.[19] Still, it is instructive to see how this particular choice impacts the results.

Therefore, we carry out additional tests using the midquote 30 minutes after the opening (at 10:00 a.m.) as an alternative definition of the end of the initial price-adjustment period. Table 2.13 and Table 2.14 compare the results of the original and the alternative definitions side by side. The first two columns (of both panels) simply show the results from Table 2.10—Table 2.12, where the initial price-adjustment period is defined as the close-to-open return and we use for the subsequent price-adjustment period the open-to-10:30 a.m. return. The last two columns use the alternative definition, whereby the initial price-adjustment period is close-to-10:00 a.m. and the subsequent price-adjustment period is 10:00 a.m.-to-10:30 a.m. Notice that both definitions begin with the closing price on day $t-1$ and end with the price at 10:30 a.m. on day t. The only difference is the end of the initial price-adjustment period: opening of the market (left two columns) versus 10:00 a.m. (right two columns).

The results in the tables are consistent with the price pattern we observe in Figure 2.1. In particular, we see that the mean and median magnitudes of DiffRetChg in categories 6 and 7 are about 50% to 100% larger when we allow more time for the initial adjustment of prices. The mean magnitude is also twice as large for the positive earnings announcements in categories 1 and 2. While the medians are not larger in categories 1 and 2, all results appear to be more statistically significant when we use the midquote 30 minutes after the open to represent the initial adjustment of prices. These results are further consistent with our model, and demonstrate that our conclusions on the manner in which influence affects the evolution of prices are not very sensitive to the various choices we need to make when bringing the model to the data.

2.4.3 Volume

One of the interesting insights that come out of our theory is the dynamic way in which the influence of published opinions by experts and media pundits affects investors. Investors

[19] Prices could also be adjusting in after-hours trading, in which case our use of the opening price in the empirical work is not necessarily equivalent to the first period in the model.

who do not observe the published view are taken aback by the fast adjustment of prices (and the illiquidity associated with it) in the initial price-adjustment period and abstain from trading. Only after prices adjust to the published view does the market become liquid enough to facilitate trading by both types of investors (those who were influenced by the published view and those who were not). As such, there is lower volume in the influence model before prices adjust to the published view, and volume increases when the wisdom of the crowds effect emerges. This volume pattern does not arise in the benchmark model.

Testing this effect is somewhat more nuanced than testing the return implications. When dealing with returns, the risk premium over a very short interval is not an important consideration relative to the movement of prices following news, and hence one can easily compare periods of varied lengths. With volume, on the other hand, increasing the length of a period over which we measure the effect necessarily means that volume in the longer period will be at least as large as in the shorter period. With this caveat in mind, though, it could be still meaningful to compare the volume patterns in afternoon and morning announcements for a given definition of initial and subsequent price-adjustment periods.

We therefore define VolChg to be volume during the subsequent price-adjustment period minus volume during the initial price-adjustment period, and empirically test:

$$\text{DiffVolChg} = \text{VolChg(Afternoon)} - \text{VolChg(Morning)} > 0, \tag{2.8}$$

pooling together categories 1, 2, 6, and 7. As the volume period for the initial price-adjustment period we follow the same definition as in Section 2.4.1: we accumulate volume starting from 6:00 p.m. on the previous day and up to and including the opening trade (or opening auction). For volume in the subsequent price-adjustment period we use two measures for robustness: (i) cumulative volume from 10:00 a.m. to 10:30 a.m. (VOL1), and (ii) cumulative volume from 10:30 a.m. to 11:00 a.m. (VOL2).[20] We also use two alternative definitions of volume. The first definition is turnover, which is the number of shares traded from TAQ divided by the number of shares outstanding from the CRSP database.[21] The second definition is dollar volume.

---

[20] Our conclusions are similar if we use periods later in the trading day. Also, the results hold if we run categories 1 and 2 separately from categories 6 and 7.

[21] Our use of turnover is motivated by Lo and Wang (2000), who discuss the theoretical advantages of using turnover as a measure of volume.

Table 2.15 shows that the means and medians of DiffVolChg are positive and statistically significant for both VOL1 and VOL2 using either turnover or dollar volume.[22] These results are consistent with the predictions of our theory, and in particular, with the idea that published views create not just price patterns but also patterns in the trading process itself.

2.5 Conclusions

When news about a firm is made public, it is not the case that the information is simply announced to the world of investors and nothing further is ever said about it. Rather, the news is picked up by media pundits, newsletter writers, on-line discussion group leaders, and other experts who present their interpretations of the news. While it is likely that they are better than the average investor in terms of interpreting the news, their interpretations are definitely not perfect. At the same time, they wield influence and hence their opinion affects the trading of other investors who listen to media reports or read the opinions in on-line newsletters and discussion groups. How does their influence impact the evolution of prices? Our goal in this paper is to answer this question first by constructing a theoretical model and second by providing some empirical work to test and calibrate the model.

We note from the outset that the idea that published opinions could introduce a shared error into prices is not new. Treynor (1987) made the point that price efficiency in the market comes from the independent opinions of a large number of investors who err independently. Even if the accuracy of the published opinion is greater than the accuracy of the the average individual opinion, an error in the published opinion will be reflected in the trading of all investors who are influenced by it, and hence would affect prices much more than an error in one individual's opinion. Our paper's contribution is twofold. First, we model this intuition rigorously using a sequential trade framework, and investigate the consequences of this idea for the evolution of prices as well as for the trading process itself. Second, we take the model to the data to see if we can find evidence consistent with the model, and then use the empirical estimates to get a sense of the magnitude of influence in in U.S. equity markets.

---

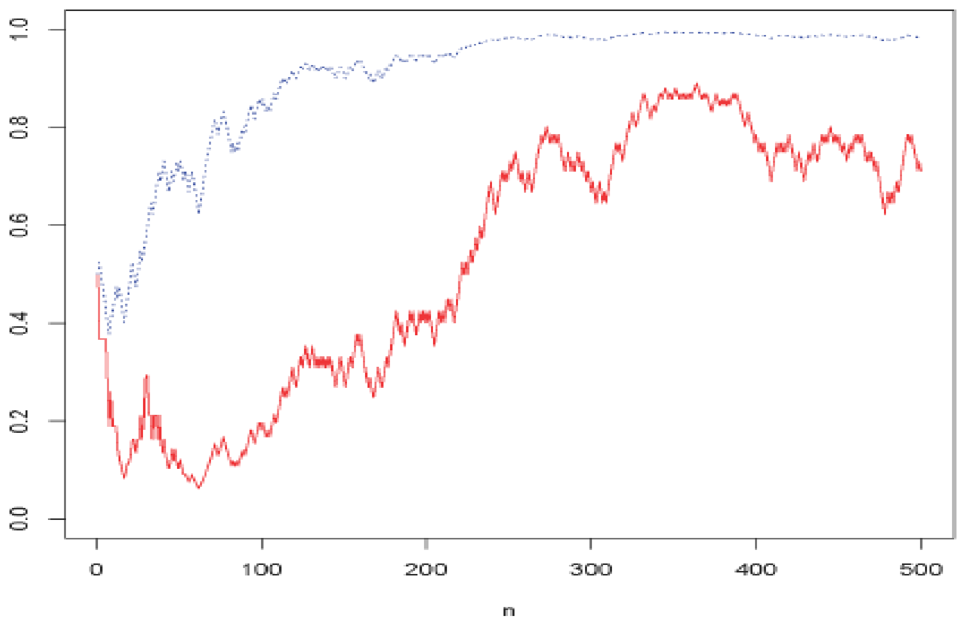
[22] Jiang, Likitapiwat, and McInish (2012) note that volume numbers may be different across the two primary listing exchanges (NYSE and NASDAQ) due to differences in market structure. If there is a systematic bias in exchange listing between firms that announce their earnings in the morning versus those that announce in the afternoon, our volume tests could be affected. To ensure that this was not the case, we also carried out the tests only on pairs for which both the morning and afternoon announcements belong to firms that are listed on the same primary exchange. The results were similar.

Our model provides novel implications pertaining to how influence matters for prices. We show that influence will initially increase the price impact associated with news, but later could slow the adjustment of prices to the true value. In the presence of influence, some investors choose to stay on the sideline during the initial price-adjustment period, reducing the intensity of trading. As prices adjust to the published view and trading costs decline, additional investors join the trading process and a wisdom of the crowds effect emerges whereby the independent views of many investors contribute to the informational efficiency of prices. Their trading can correct the shared error that was introduced by the published opinion, but at the cost of slower convergence of prices to the true value compared with what would happen in an economy where a published view does not yield influence.

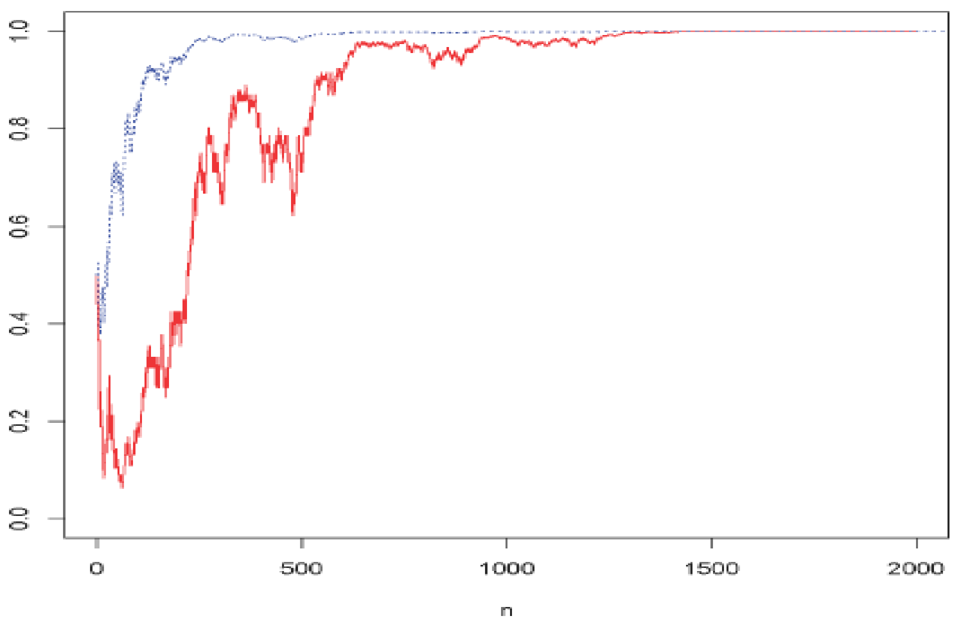
Unlike in information cascades papers, the arrival of investors in our model changes the terms of trade for subsequent investors. As such, the decision rule changes over time endogenously, and the error does not propagate forever. We do not need exogenous arrival of public information or investors with high-precision information that arrive late in the sequence of trading to counter the impact of the error (see, for example, Bikhchandani, Hirshleifer, and Welch (1992)). Rather, as arriving investors face evolving prices and trading costs over time, the population of investors who choose to trade changes, leading to the elimination of the error by letting the wisdom of the crowds take effect.

Following Treynor's observation that the number of influenced investors increases with the time elapsed since the release of the published opinion, we use the timing of earnings announcements (or the interval of time between the announcement and the opening of the next regular trading session) to test our theory. We find evidence consistent with the predictions of the model, and the results are robust to various choices we make in the empirical specifications of the tests.

We hope that future work will use the insights of our model to assess how the influence of media impacts prices in other markets or under different circumstances. As the plurality of media forms and the prevalence of media in our lives increase over time, the influence media exerts over financial markets remains an important topic for study.



(a) Panel A: Price Path in the Influence and Benchmark Models up to Period 500



(b) Panel B: Price Path in the Influence and Benchmark Models up to Period 3,000

Figure 2.1: Price Adjustment when the Published View Is Incorrect

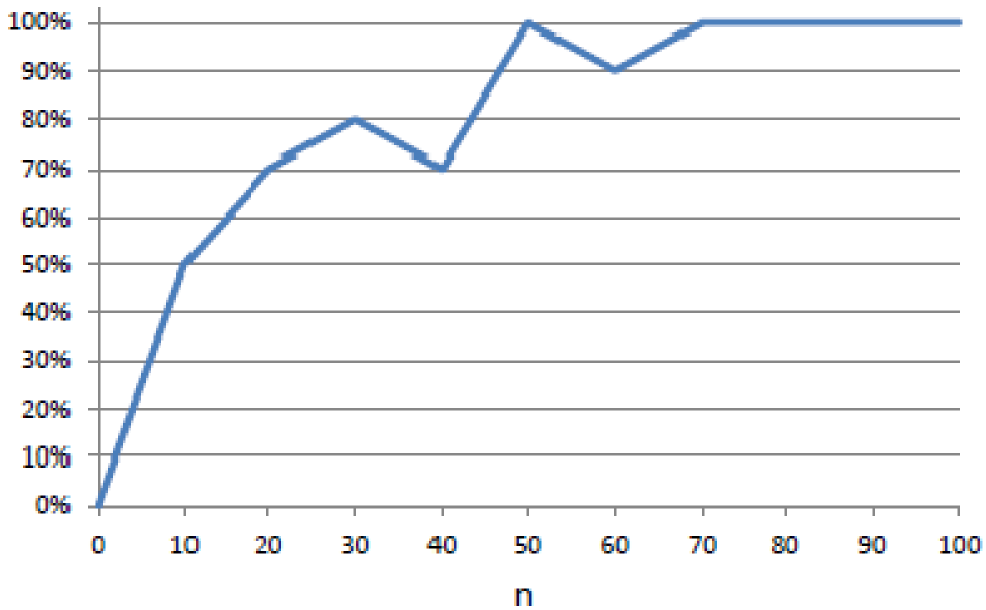


Figure 2.2: Volume in the Influence Model

Table 2.1: Trader Type Definitions

Type (θ)	$\tilde{v}^e$	$\tilde{v}^n$	Description
1	1	1	Positive published view; Positive personal view
2	1	0	Positive published view; Negative personal view
3	N/A	1	Positive personal view
4	N/A	0	Negative personal view
5	0	1	Negative published view; Positive personal view
6	0	0	Negative published view; Negative personal view
7	N/A	N/A	Noise Trader Buyer
8	N/A	N/A	Noise Trader Seller

Table 2.2: Trader Type Valuations

Type (θ)	$E[\tilde{v} \mid H, \tilde{\theta} = \theta]$
1	$\frac{\pi p_n^h}{\pi p_n^h + (1-\pi)(1-p_n^h)}$
2	$\frac{(1-\pi) p_n^h}{(1-\pi) p_n^h + \pi(1-p_n^h)}$
3	$\frac{\pi p_n^m}{\pi p_n^m + (1-\pi)(1-p_n^m)}$
4	$\frac{(1-\pi) p_n^m}{(1-\pi) p_n^m + \pi(1-p_n^m)}$
5	$\frac{\pi p_n^l}{\pi p_n^l + (1-\pi)(1-p_n^l)}$
6	$\frac{(1-\pi) p_n^l}{(1-\pi) p_n^l + \pi(1-p_n^l)}$
7	p_n^m
8	p_n^m

Table 2.3: Number of Earnings Announcements by Year and Earnings Surprise Category

	Years	2000	2001	2002	2003	2004	2005
Entire Sample	Morning	2,961	3,634	4,077	4,516	4,954	4,866
	Afternoon	1,921	2,944	3,411	3,966	4,714	4,892
	Total	4,882	6,578	7,488	8,482	9,668	9,758
Category 1	Morning	496	600	772	895	1,017	973
	Afternoon	378	572	712	886	981	1,224
	Total	874	1,172	1,484	1,781	1,998	2,197
Category 2	Morning	560	589	749	845	1,002	1,008
	Afternoon	469	454	664	884	1,023	1,259
	Total	1,029	1,043	1,413	1,729	2,025	2,267
Category 6	Morning	247	399	342	381	494	541
	Afternoon	198	351	276	422	524	633
	Total	445	750	618	803	1,018	1,174
Category 7	Morning	264	412	362	452	418	496
	Afternoon	224	354	382	417	406	621
	Total	488	766	744	869	824	1,117

Table 2.4: Number of Earnings Announcement Pairs by Industry

Industry	Category 1	Category 2	Category 6	Category 7	Sample
Consumer Nondurables	239	292	140	82	1547
Consumer Durables	155	169	81	89	794
Manufacturing	577	716	359	340	3591
Oil, Gas, and Coal Extraction and Products	221	232	127	80	1037
Business Equipment	869	689	329	411	3571
Telephone and Television Transmission	170	154	76	125	721
Wholesale, Retail, and Some Services	486	582	239	251	3079
Healthcare, Medical Equipment, and Drugs	632	416	252	330	2706
Utilities	242	185	136	95	942
Other	1162	1318	665	601	7020

Table 2.5: Summary Statistics: Categories 1&2 Matching Using Industry Classification, Market Capitalization, and Price

		10%	25%	50%	75%	90%
MktCap ($millions)	Morning	113.0	257.5	706.0	2,359.00	8,225.50
	Afternoon	116.7	256.5	703.9	2,326.20	7,354.30
Price ($)	Morning	5.8	10.8	19.5	31.7	45.7
	Afternoon	5.9	11	19.5	32.5	46.3
Earnings Surprise	Morning	0.0008	0.0012	0.0022	0.0048	0.0120
	Afternoon	0.0009	0.0013	0.0025	0.0049	0.0108

Table 2.6: Summary Statistics: Categories 1&2 Matching Using Market Capitalization and Price

		10%	25%	50%	75%	90%
MktCap ($millions)	Morning	113.0	257.5	706.0	2,359.00	8,225.50
	Afternoon	112.8	258.2	703.2	2,350.60	8,068.40
Price ($)	Morning	5.8	10.8	19.5	31.7	45.7
	Afternoon	5.8	10.8	19.5	31.9	45.5
Earnings Surprise	Morning	0.0008	0.0012	0.0022	0.0048	0.0120
	Afternoon	0.001	0.0013	0.0025	0.005	0.0114

Table 2.7: Summary Statistics: Categories 6&7 Matching Using Industry Classification, Market Capitalization, and Price

		10%	25%	50%	75%	90%
MktCap ($millions)	Morning	70.8	158.8	431.1	1,271.7	3,890.0
	Afternoon	72.2	158.3	436.4	1,222.8	3,662.7
Price ($)	Morning	3.8	6.9	13.6	23.4	35.9
	Afternoon	3.9	7.0	13.6	23.5	35.0
Earnings Surprise	Morning	-0.0259	-0.0117	-0.0045	-0.0021	-0.0014
	Afternoon	-0.0343	-0.0131	-0.0054	-0.0025	-0.0016

Table 2.8: Summary Statistics: Categories 6&7 Matching Using Market Capitalization and Price

		10%	25%	50%	75%	90%
MktCap ($millions)	Morning	70.8	158.8	431.1	1,271.7	3,890.0
	Afternoon	70.9	158.6	434.2	1,277.4	3,863.4
Price ($)	Morning	3.8	6.9	13.6	23.4	35.9
	Afternoon	3.8	6.9	13.6	23.4	35.6
Earnings Surprise	Morning	-0.0259	-0.0117	-0.0045	-0.0021	-0.0014
	Afternoon	-0.0343	-0.0129	-0.0054	-0.0023	-0.0016

Table 2.9: Initial Price Adjustment

	Categories 1&2		Categories 6&7	
	Mean Diff	Median Diff	Mean Diff	Median Diff
AbsRet	0.0096	0.003	0.0119	0.0023
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Amihud's Measure	0.0962	0.0024	0.1304	0.0053
	(<0.001)	(<0.001)	(<0.001)	(<0.001)

Table 2.10: Return Patterns: Raw Returns

	Categories 1&2		Categories 6&7	
	Mean DiffRetChg	Median DiffRetChg	Mean DiffRetChg	Median DiffRetChg
Close-to-10:30am	-0.0023	-0.0023	0.0057	0.0028
	(0.007)	(<0.001)	(<0.001)	(<0.001)
Close-to-11:00am	-0.0021	-0.0019	0.0045	0.0028
	(0.016)	(<0.001)	(0.002)	(<0.001)
Close-to-11:30am	-0.0021	-0.0018	0.0047	0.0038
	(0.018)	(0.001)	(0.001)	(<0.001)
Close-to-12:00pm	-0.0024	-0.0021	0.0044	0.0035
	(0.008)	(<0.001)	(0.003)	(<0.001)

Table 2.11: Return Patterns: Returns in Excess of the S&P 500

	Categories 1&2		Categories 6&7	
	Mean DiffRetChg	Median DiffRetChg	Mean DiffRetChg	Median DiffRetChg
Close-to-10:30am	-0.0022	-0.0022	0.0058	0.0031
	(0.010)	(<0.001)	(<0.001)	(<0.001)
Close-to-11:00am	-0.0018	-0.0021	0.0043	0.0030
	(0.031)	(0.002)	(0.003)	(0.001)
Close-to-11:30am	-0.002	-0.0019	0.0045	0.0030
	(0.021)	(0.001)	(0.002)	(0.001)
Close-to-12:00pm	-0.0022	-0.0017	0.0042	0.0036
	(0.012)	(0.001)	(0.005)	(0.001)

Table 2.12: Return Patterns: Abnormal Returns Relative to Preannouncement Period

	Categories 1&2		Categories 6&7	
	Mean DiffRetChg	Median DiffRetChg	Mean DiffRetChg	Median DiffRetChg
Close-to-10:30am	-0.002	-0.0019	0.0057	0.0032
	(0.024)	(0.001)	(<0.001)	(<0.001)
Close-to-11:00am	-0.0019	-0.0019	0.0042	0.0031
	(0.035)	(0.003)	(0.005)	(<0.001)
Close-to-11:30am	-0.002	-0.0019	0.0045	0.0029
	(0.028)	(0.003)	(0.003)	(<0.001)
Close-to-12:00pm	-0.0022	-0.0015	0.0042	0.0037
	(0.018)	(0.002)	(0.007)	(<0.001)

Table 2.13: Return Patterns with Alternative Initial Adjustment Period: Categories 1&2 (Positive Earnings Surprise)

	Close-to-Open Initial Price Adjustment		Close-to-10:00am Initial Price Adjustment	
	Mean DiffRetChg	Median DiffRetChg	Mean DiffRetChg	Median DiffRetChg
Raw Return	-0.0023	-0.0023	-0.0047	-0.0023
	(0.007)	(<0.001)	(<0.001)	(<0.001)
Excess Return	-0.0022	-0.0022	-0.0046	-0.0022
	(0.010)	(<0.001)	(<0.001)	(<0.001)
Abnormal Return	-0.0022	-0.0015	-0.0054	-0.0031
	(0.018)	(0.002)	(<0.001)	(<0.001)

Table 2.14: Return Patterns with Alternative Initial Adjustment Period: Categories 6&7 (Negative Earnings Surprise)

	Close-to-Open Initial Price Adjustment		Close-to-10:00am Initial Price Adjustment	
	Mean DiffRetChg	Median DiffRetChg	Mean DiffRetChg	Median DiffRetChg
Raw Return	0.0057	0.0028	0.0100	0.0046
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Excess Return	0.0058	0.0031	0.0099	0.0048
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Abnormal Return	0.0057	0.0032	0.0086	0.0052
	(<0.001)	(<0.001)	(<0.001)	(<0.001)

Table 2.15: Volume

	% Turnover		$ Volume (1,000)	
	Mean DiffVolChg	Median DiffVolChg	Mean DiffVolChg	Median DiffVolChg
Measure I	0.07	0.03	3,018.20	133.4
	(<0.001)	(<0.001)	(<0.001)	(<0.001)
Measure II	0.05	0.02	2,323.70	94.5
	(<0.001)	(<0.001)	(<0.001)	(<0.001)

APPENDIX A

PROOFS IN CHAPTER 1

This appendix specifies how to calculate the conditional and posterior probabilities and proves the lemmas in Chapter 1. I assume the probabilities of an error term equalling zero for asset 1 and 2 are the same for simplification, i.e. $\pi_1 = \pi_2 = \pi$.

A.1 Conditional Probability

The conditional probabilities of a view on the terminal value of each risky asset j are easy to compute, where $j = 1, 2$. Using the relation between $\tilde{v}_j^s$ and $\tilde{v}_j$, I have the following:

$$\begin{cases} \pi_{hhj} = Prob(v_j^s = v_h | v_j = v_h) = Prob(\epsilon_j^s = 0) = \pi \\ \pi_{lhj} = Prob(v_j^s = v_l | v_j = v_h) = Prob(\epsilon_j^s = 1) = 1 - \pi \\ \pi_{hlj} = Prob(v_j^s = v_h | v_j = v_l) = Prob(\epsilon_j^s = 1) = 1 - \pi \\ \pi_{llj} = Prob(v_j^s = v_l | v_j = v_l) = Prob(\epsilon_j^s = 0) = \pi. \end{cases} \tag{A.1}$$

In terms of the index, I compute the probability for the case $\tilde{V}_I^i = V_M$ and $\tilde{V}_I = V_M$ and the other cases follow the same method.

$$\begin{aligned}
\pi_{MMI} &= Prob(V_I^i = V_M | V_I = V_M) = \frac{Prob(V_I^i = V_M, V_I = V_M)}{prob(V_I = V_M)} \\
&= \frac{Prob(V_I^i = V_M, v_1 = v_h, v_2 = v_l) + Prob(V_I^i = V_M, v_1 = v_l, v_2 = v_h)}{Prob(V_I = V_M)} \\
&= \frac{Prob(V_I^i = V_M | v_1 = v_h, v_2 = v_l) * Prob(v_1 = v_h, v_2 = v_l)}{Prob(V_I = V_M)} + \\
&\quad \frac{Prob(V_I^i = V_M | v_1 = v_l, v_2 = v_h) * Prob(v_1 = v_l, v_2 = v_h)}{Prob(V_I = V_M)} \\
&= \frac{1}{2} Prob(V_I^i = V_M | v_1 = v_h, v_2 = v_l) + \frac{1}{2} Prob(V_I^i = V_M | v_1 = v_l, v_2 = v_h) \\
&= \frac{1}{2}[Prob(v_1^i = v_h, v_2^i = v_l | v_1 = v_h, v_2 = v_l) + Prob(v_1^i = v_l, v_2^i = v_h | v_1 = v_h, v_2 = v_l)] + \\
&\quad \frac{1}{2}[Prob(v_1^i = v_h, v_2^i = v_l | v_1 = v_l, v_2 = v_h) + Prob(v_1^i = v_l, v_2^i = v_h | v_1 = v_l, v_2 = v_h)] \\
&= \frac{1}{2}[\pi^2 + (1 - \pi)^2] + \frac{1}{2}[(1 - \pi)^2 + \pi^2] \\
&= \pi^2 + (1 - \pi)^2,
\end{aligned}$$

where the third equation uses the fact that two cases $(v_1 = v_h, v_2 = v_l)$ or $(v_1 = v_l, v_2 = v_h)$ give $V_I = V_M$, the fifth equation follows from the independent assumption and the prior distribution of assets, and the seventh equation also uses the independent assumption and results from (A.1). Then following the same method, the conditional probabilities for the index are given by:

$$\begin{cases} \pi_{HHI} = Prob(V_I^i = V_H | V_I = V_H) = \pi^2 \\ \pi_{MHI} = Prob(V_I^i = V_M | V_I = V_H) = 2\pi(1-\pi) \\ \pi_{LHI} = Prob(V_I^i = V_L | V_I = V_H) = (1-\pi)^2 \\ \pi_{HMI} = Prob(V_I^i = V_H | V_I = V_M) = \pi(1-\pi) \\ \pi_{MMI} = Prob(V_I^i = V_M | V_I = V_M) = \pi^2 + (1-\pi)^2 \\ \pi_{LMI} = Prob(V_I^i = V_L | V_I = V_M) = \pi(1-\pi) \\ \pi_{HLI} = Prob(V_I^i = V_H | V_I = V_L) = (1-\pi)^2 \\ \pi_{MLI} = Prob(V_I^i = V_M | V_I = V_L) = 2\pi(1-\pi) \\ \pi_{LLI} = Prob(V_I^i = V_L | V_I = V_L) = \pi^2 \end{cases} \tag{A.2}$$

Now with (A.1) and (A.2), I can calculate the conditional probability for a specific realization of O or O'. I denote N^o defined in the main text the numbers of occurrence of relevant values for that specific realization. Due to different trading behaviours, I compute the probability for indexers and stock pickers respectively.

For indexers, I have the following when $V_I = V_H$:

$$\begin{aligned} Prob(O' | V_I = V_H) &= Prob(V_I^1 | V_I = V_H) * ... * Prob(V_I^N | V_I = V_H) \\ &= \pi_{HHI}^{(N_{Hs}+N_{Hi})} * \pi_{MHI}^{(N_{Ms}+N_{Mi})} * \pi_{LHI}^{(N_{Ls}+N_{Li})}, \end{aligned}$$

where the first equation follows from the independent assumption of the errors and the second equation just counts the relevant numbers of occurrence. Using the same methodology, I compute the probabilities for all cases:

$$\begin{cases} Prob(O' | V_I = V_H) = \pi_{HHI}^{(N_{Hs}+N_{Hi})} * \pi_{MHI}^{(N_{Ms}+N_{Mi})} * \pi_{LHI}^{(N_{Ls}+N_{Li})} \\ Prob(O' | V_I = V_M) = \pi_{HMI}^{(N_{Hs}+N_{Hi})} * \pi_{MMI}^{(N_{Ms}+N_{Mi})} * \pi_{LMI}^{(N_{Ls}+N_{Li})} \\ Prob(O' | V_I = V_L) = \pi_{HLI}^{(N_{Hs}+N_{Hi})} * \pi_{MLI}^{(N_{Ms}+N_{Mi})} * \pi_{LLI}^{(N_{Ls}+N_{Li})}, \end{cases} \tag{A.3}$$

where $N_{Hs} + N_{Ms} + N_{Ls} = (1-\alpha)N$ and $N_{Hi} + N_{Mi} + N_{Li} = \alpha N$.

In terms of stock pickers, I need to compute the probabilities for events in $\{(v_1 = v_h, v_2 = v_h), (v_1 = v_h, v_2 = v_l), (v_1 = v_l, v_2 = v_h), (v_1 = v_l, v_2 = v_l)\}$. For event $(v_1 = v_h, v_2 = v_h)$, I have:

$$
\begin{aligned}
&Prob(O|v_1 = v_h, v_2 = v_h) \\
&= Prob(v_1^1, v_2^1|v_1 = v_h, v_2 = v_h) * ... * Prob(v_1^{(1-\alpha)N}, v_2^{(1-\alpha)N}|v_1 = v_h, v_2 = v_h) * \\
&\quad Prob(V_I^1|v_1 = v_h, v_2 = v_h) * ... * Prob(V_I^{\alpha N}|v_1 = v_h, v_2 = v_h) \\
&= \pi_{hh1}^{N_{h1}} * \pi_{lh1}^{N_{l1}} * \pi_{hh2}^{N_{h2}} * \pi_{hl2}^{N_{l2}} * Prob(V_I^i = V_H|v_1 = v_h, v_2 = v_h)^{N_{Hi}} * \\
&\quad Prob(V_I^i = V_L|v_1 = v_h, v_2 = v_h)^{N_{Li}} * Prob(V_I^i = V_M|v_1 = v_h, v_2 = v_h)^{N_{Mi}} \\
&= \pi_{hh1}^{(N_{h1}+N_{Hi})} * \pi_{lh1}^{(N_{l1}+N_{Li})} * \pi_{hh2}^{(N_{h2}+N_{Hi})} * \pi_{lh2}^{(N_{l2}+N_{Li})} * \\
&\quad \sum_{z=0}^{N_{Mi}} [\pi_{hh1}^{z} * \pi_{lh1}^{(N_{Mi}-z)} * \pi_{hh2}^{(N_{Mi}-z)} * \pi_{lh2}^{z}],
\end{aligned}
$$

where z denotes the possible number of occurrence of $v_1^i = v_h$ when V_M occurs N_{Mi} times. Equations one and two use the independent assumption among error terms and risky assets. The third equation follows from two facts. The first fact is that by knowing $V_I^i = V_H$ or $V_I^i = V_L$, a stock picker tells exactly the indexer's opinion for each individual asset. The second is that a stock picker can not tell the exact opinions for individual assets by seeing $V_I^i = V_M$ so he needs to count all possibilities that give N_{Mi} times of V_M. Following the above method, I have:

$$
\begin{cases}
Prob(O|v_1 = v_h, v_2 = v_h) = & \pi_{hh1}^{(N_{h1}+N_{Hi})} * \pi_{lh1}^{(N_{l1}+N_{Li})} * \pi_{hh2}^{(N_{h2}+N_{Hi})} * \pi_{lh2}^{(N_{l2}+N_{Li})} * \\
& \sum_{z=0}^{N_{Mi}} [\pi_{hh1}^{z} * \pi_{lh1}^{(N_{Mi}-z)} * \pi_{hh2}^{(N_{Mi}-z)} * \pi_{lh2}^{z}] \\
Prob(O|v_1 = v_h, v_2 = v_l) = & \pi_{hh1}^{(N_{h1}+N_{Hi})} * \pi_{lh1}^{(N_{l1}+N_{Li})} * \pi_{hl2}^{(N_{h2}+N_{Hi})} * \pi_{ll2}^{(N_{l2}+N_{Li})} * \\
& \sum_{z=0}^{N_{Mi}} [\pi_{hh1}^{z} * \pi_{lh1}^{(N_{Mi}-z)} * \pi_{hl2}^{(N_{Mi}-z)} * \pi_{ll2}^{z}] \\
Prob(O|v_1 = v_l, v_2 = v_h) = & \pi_{hl1}^{(N_{h1}+N_{Hi})} * \pi_{ll1}^{(N_{l1}+N_{Li})} * \pi_{hh2}^{(N_{h2}+N_{Hi})} * \pi_{lh2}^{(N_{l2}+N_{Li})} * \\
& \sum_{z=0}^{N_{Mi}} [\pi_{hl1}^{z} * \pi_{ll1}^{(N_{Mi}-z)} * \pi_{hh2}^{(N_{Mi}-z)} * \pi_{lh2}^{z}] \\
Prob(O|v_1 = v_l, v_2 = v_l) = & \pi_{hl1}^{(N_{h1}+N_{Hi})} * \pi_{ll1}^{(N_{l1}+N_{Li})} * \pi_{hl2}^{(N_{h2}+N_{Hi})} * \pi_{ll2}^{(N_{l2}+N_{Li})} * \\
& \sum_{z=0}^{N_{Mi}} [\pi_{hl1}^{z} * \pi_{ll1}^{(N_{Mi}-z)} * \pi_{hl2}^{(N_{Mi}-z)} * \pi_{ll2}^{z}],
\end{cases}
\tag{A.4}
$$

where $N_{h1} + N_{l1} = (1-\alpha)N$ and $N_{h2} + N_{l2} = (1-\alpha)N$.

A.2 Proof of Lemmas and Posterior Probability

Proof of Lemma 1. Since $I(X = i)$ is an indicator function, then the probability mass function of X is written as:

$$p(X;\mathbf{p}) = \prod_{i=0}^{k} p_i^{I(X=i)}.$$

So we can write the joint probability distribution of the independent sample:

$$p(X_1, X_2, ..., X_N; \mathbf{p}) = \prod_{i=0}^{k} p_i^{N_i}, \tag{A.5}$$

where $N_i = \sum_{n=1}^{N} I(X_n = i)$.

Moreover, $\hat{\mathbf{p}}$ is a multinomial variable since it represents the number of occurrence of i in N independent trials of X. We can write its probability distribution as:

$$p(N_0, N_1, N_2, ..., N_k; \mathbf{p}) = \binom{N}{N_0, ..., N_k} \prod_{i=0}^{k} p_i^{N_i}$$

Now we could write (A.5) this way:

$$\begin{aligned} p(X_1, X_2, ..., X_N; \mathbf{p}) &= \binom{N}{N_0, ..., N_k}^{-1} p(N_0, N_1, N_2, ..., N_k; \mathbf{p}) \\ &= \binom{N}{N_0, ..., N_k}^{-1} p(\hat{\mathbf{p}}; \mathbf{p}), \end{aligned}$$

and this shows that $\hat{\mathbf{p}} = \{N_0, N_1, N_2, ..., N_k\}$ is a sufficient statistic for $\mathbf{p}$.

Proof of Lemma 2. I prove the lemma for stock pickers and event $(v_1 = v_h, v_2 = v_h)$ only. The other events and the case of indexers follow the same procedure.

By Lemma 1, $Prob(v_1^s, v_2^s, O|v_1 = v_h, v_2 = v_h) = Prob(v_1^s, v_2^s, N^o|v_1 = v_h, v_2 = v_h)$. Then I only need to discuss different realizations of (v_1^s, v_2^s). When $(v_1^s = v_h, v_2^s = v_h)$, I have the following:

$$\begin{aligned} &Prob(v_1^s, v_2^s, N^o|v_1 = v_h, v_2 = v_h) \\ &= Prob(v_1^s = v_h, v_2^s = v_h, (N_{h1} - 1), N_{l1}, (N_{h2} - 1), N_{l2}, N_{Hi}, N_{Mi}, N_{Li}|v_1 = v_h, v_2 = v_h) \\ &= \pi_{hh1} * \pi_{hh1}^{(N_{h1}-1+N_{Hi})} * \pi_{lh1}^{(N_{l1}+N_{Li})} * \pi_{hh2} * \pi_{hh2}^{(N_{h2}-1+N_{Hi})} * \pi_{lh2}^{(N_{l2}+N_{Li})} * \\ &\quad \sum_{z=0}^{N_{Mi}} [\pi_{hh1}^{z} * \pi_{lh1}^{(N_{Mi}-z)} * \pi_{hh2}^{(N_{Mi}-z)} * \pi_{lh2}^{z}] \\ &= \pi_{hh1}^{(N_{h1}+N_{Hi})} * \pi_{lh1}^{(N_{l1}+N_{Li})} * \pi_{hh2}^{(N_{h2}+N_{Hi})} * \pi_{lh2}^{(N_{l2}+N_{Li})} * \sum_{z=0}^{N_{Mi}} [\pi_{hh1}^{z} * \pi_{lh1}^{(N_{Mi}-z)} * \pi_{hh2}^{(N_{Mi}-z)} * \pi_{lh2}^{z}] \\ &= Prob(O|v_1 = v_h, v_2 = v_h). \end{aligned}$$

For the other realizations of (v_1^s, v_2^s), they all equal $Prob(O|v_1 = v_h, v_2 = v_h)$. Then following this way, I can show results for the rest events and also for indexers.

By the above lemmas, I have the following for indexers in the case of $V_I = V_H$:

$$
\begin{aligned}
&Prob(V_I = V_H | V_I^i, P_1(O), P_2(O)) = Prob(V_I = V_H | N^o) \\
&= \frac{Prob(V_I = V_H) * Prob(N^o | V_I = V_H)}{\sum_X Prob(V_I) * Prob(N^o | V_I)},
\end{aligned}
\tag{A.6}
$$

where $X = \{V_H, V_M, V_L\}$. Next using the corresponding conditional probability in (A.3), I can compute the posterior. The other cases follow the same procedure.

For stock pickers, I first calculate the posterior for each event. When $(v_1 = v_h, v_2 = v_h)$, I have:

$$
\begin{aligned}
&Prob(v_1 = v_h, v_2 = v_h | v_1^s, v_2^s, P_1(O), P_2(O)) = Prob(v_1 = v_h, v_2 = v_h | N^o) \\
&= \frac{Prob(v_1 = v_h, v_2 = v_h) * Prob(N^o | v_1 = v_h, v_2 = v_h)}{\sum_e Prob(v_1, v_2) * Prob(N^o | v_1, v_2)},
\end{aligned}
$$

where e in $\{(v_1 = v_h, v_2 = v_h), (v_1 = v_h, v_2 = v_l), (v_1 = v_l, v_2 = v_h), (v_1 = v_l, v_2 = v_l)\}$. Using the results in (A.4), I can compute the posterior. Then the rest events follow.

Now with the posteriors of events, I can calculate the posterior probabilities for each risky asset. For asset 1, I have:

$$
\begin{aligned}
&Prob(v_1 = v_h | v_1^s, v_2^s, P_1(O), P_2(O)) \\
&= Prob(v_1 = v_h, v_2 = v_h | v_1^s, v_2^s, P_1(O), P_2(O)) + Prob(v_1 = v_h, v_2 = v_l | v_1^s, v_2^s, P_1(O), P_2(O)).
\end{aligned}
\tag{A.7}
$$

Then the posterior of $v_1 = v_l$ is given by:

$$
Prob(v_1 = v_l | v_1^s, v_2^s, P_1(O), P_2(O)) = 1 - Prob(v_1 = v_h | v_1^s, v_2^s, P_1(O), P_2(O)). \tag{A.8}
$$

And the posteriors of asset 2 follow the same method.

APPENDIX B

PROOFS IN CHAPTER 2

This appendix provides proofs for the propositions shown in Chapter 2.

Lemma 3 *Given a subset of types, Θ, and an arbitrary event E, we have:*

$$E[\tilde{v}|\tilde{\theta} \in \Theta, E] = \frac{\sum_{\theta \in \Theta} E[\tilde{v}|E, \tilde{\theta} = \theta] P(\tilde{\theta} = \theta|E)}{\sum_{\theta \in \Theta} P(\tilde{\theta} = \theta|E)}. \tag{B.1}$$

Proof of Lemma 3. Because $\tilde{v}$ can only take on the values zero or one, we have:

$$\begin{aligned} E[\tilde{v}|E, \tilde{\theta} \in \Theta] &= P(\tilde{v} = 1|E, \tilde{\theta} \in \Theta) \\ &= \frac{P(\tilde{v} = 1, \tilde{\theta} \in \Theta|E)}{P(\tilde{\theta} \in \Theta|E)} \\ &= \frac{\sum_{\theta \in \Theta} P(\tilde{v} = 1, \tilde{\theta} = \theta|E)}{P(\tilde{\theta} \in \Theta|E)} \\ &= \frac{\sum_{\theta \in \Theta} P(\tilde{v} = 1|\tilde{\theta} = \theta|E) P(\tilde{\theta} = \theta|E)}{P(\tilde{\theta} \in \Theta|E)} \\ &= \frac{\sum_{\theta \in \Theta} E[\tilde{v} = 1|\tilde{\theta} = \theta|E] P(\tilde{\theta} = \theta|E)}{P(\tilde{\theta} \in \Theta|E)}. \end{aligned}$$

Proof of Proposition 2. Let the state variables be given as in (2.4). Our proof proceeds in two steps. First, we assume that each of the three state variables is in the open interval $(0, 1)$. We then show that, given the state variables, we can compute the equilibrium in a given period. In the second step we show how the state variables are updated between periods. In particular, each of the state variables remains in the open interval $(0, 1)$. For brevity, we omit the period subscript n, and use the notation $\mathbf{p} = (p^l, p^m, p^h)$.

Before trade commences, one of four events is realized. The events and their probabilities (conditional on the history) in terms of the state variables are reported in Table B.1.

Table 2.1 in the paper lists the eight trader types in the model, and Table 2.2 reports the values each trader type assigns to the asset (i.e., the conditional expectations $E[\tilde{v}|H, \tilde{\theta}]$). Table B.2 below presents the probabilities that each type arrives in the market in terms of the state variables.

To compute the equilibrium we need to solve the following four equations:

$$
\begin{aligned}
B &= \{\theta : 1 \leq \theta \leq 6\ ,\ ask \leq E[\tilde{v}|H, \tilde{\theta} = \theta]\} \cup \{7\} \\
S &= \{\theta : 1 \leq \theta \leq 6\ ,\ E[\tilde{v}|H, \tilde{\theta} = \theta] \leq bid\} \cup \{8\} \\
ask = E[\tilde{v}|H, \tilde{\theta} \in B] &= \frac{\sum_{\theta \in B} E[\tilde{v}|H, \tilde{\theta} = \theta] P(\tilde{\theta} = \theta|H)}{\sum_{\theta \in B} P(\tilde{\theta} = \theta|H)} \\
bid = E[\tilde{v}|H, \tilde{\theta} \in S] &= \frac{\sum_{\theta \in S} E[\tilde{v}|H, \tilde{\theta} = \theta] P(\tilde{\theta} = \theta|H)}{\sum_{\theta \in S} P(\tilde{\theta} = \theta|H)},
\end{aligned}
\tag{B.2}
$$

where B and S are the sets of types that buy and sell, respectively, and we use Lemma 3 to express the ask and bid prices as weighted averages. Given the equilibrium ask and bid prices, we define the set of types that abstain from trading as:

$$
A = \{\theta : 1 \leq \theta \leq 6\ ,\ bid < E[\tilde{v}|H, \tilde{\theta} = \theta] < ask\}.
$$

Using the entries in Table B.2 and Table 2.2, we reduce the system of equations (B.2) that defines the equilibrium to a system of equations that depends solely on the three state variables. To see that the system (B.2) has a solution, we rank the types of investors by their valuations (the first six entries in Table 2.2). To find the ask price, we start with a guess that set B contains all six types of investors (plus the noise trader who buys), and compute the asking price that corresponds to the guess. We then check if indeed the valuation of each type of investor in B is greater than or equal to the resulting asking price. If not, we remove the investor type with the lowest valuation from the set and repeat. Because we always have at least one type with valuation strictly greater than p^m (namely, $\theta = 1$), this iterative process must terminate successfully. Moreover, the process can never terminate before we remove all types of investors with valuations strictly smaller than p^m. This proves that the equilibrium ask price is strictly greater than p^m. Similarly, we find the bid and see that it is strictly smaller than p^m. Therefore, the bid-ask spread is strictly positive.

So far we have shown how to compute the equilibrium in a given period as a function of the state variables. Denote the solution of the system (B.2) by $B(\mathbf{p})$, $A(\mathbf{p})$, $S(\mathbf{p})$, $ask(\mathbf{p})$, and $bid(\mathbf{p})$. To complete the description of the Markovian equilibrium, we need to show how the state variables evolve as the history of trade unfolds. The initial values of the state variables (i.e., before trading commences) are:

$$
\begin{aligned}
p^h &= \pi_e \\
p^m &= 1/2 \\
p^l &= (1 - \pi_e).
\end{aligned}
$$

Let $h \in \{buy, abstain, sell\}$ and $H' = \{H, h\}$. Our goal is to compute:

$$p^{h\prime} = P(\tilde{v} = 1 | H', \tilde{v}^e = 1)$$
$$p^{m\prime} = P(\tilde{v} = 1 | H')$$
$$p^{l\prime} = P(\tilde{v} = 1 | H', \tilde{v}^e = 0)$$

in terms of $\mathbf{p}$ and h. Let

$$\Theta(h, \mathbf{p}) = \begin{cases} B(\mathbf{p}) & h = buy \\ A(\mathbf{p}) & h = abstain \\ S(\mathbf{p}) & h = sell. \end{cases}$$

The posterior $p^{h\prime}$ is:

$$p^{h\prime} = E[\tilde{v} | H', \tilde{v}^e = 1] = \frac{\sum_{\theta \in \Theta(h,\mathbf{p})} E[\tilde{v} | H, \tilde{v}^e = 1, \tilde{\theta} = \theta] P(\tilde{\theta} = \theta | H, \tilde{v}^e = 1)}{\sum_{\theta \in \Theta(h,\mathbf{p})} P(\tilde{\theta} = \theta | H, \tilde{v}^e = 1)}, \quad \text{(B.3)}$$

where the terms $P(\tilde{\theta} = \theta | H, \tilde{v}^e = 1)$ and $E[\tilde{v} | H, \tilde{v}^e = 1, \tilde{\theta} = \theta]$ are given, in terms of the state variables, in Tables B.3 and B.4.

The posterior $p^{m\prime}$ is:

$$p^{m\prime} = E[\tilde{v} | H'] = \frac{\sum_{\theta \in \Theta(h,\mathbf{p})} E[\tilde{v} | H, \tilde{\theta} = \theta] P(\tilde{\theta} = \theta | H)}{\sum_{\theta \in \Theta(h,\mathbf{p})} P(\tilde{\theta} = \theta | H)}, \quad \text{(B.4)}$$

where the terms $P(\tilde{\theta} = \theta | H)$ and $E[\tilde{v} | H, \tilde{\theta} = \theta]$ are given, in terms of the state variables, in Table B.2 and Table 2.2.

Finally, the posterior $p^{l\prime}$ is:

$$p^{l'} = E[\tilde{v} | H', \tilde{v}^e = 0] = \frac{\sum_{\theta \in \Theta(h,\mathbf{p})} E[\tilde{v} | H, \tilde{v}^e = 0, \tilde{\theta} = \theta] P(\tilde{\theta} = \theta | H, \tilde{v}^e = 0)}{\sum_{\theta \in \Theta(h,\mathbf{p})} P(\tilde{\theta} = \theta | H, \tilde{v}^e = 0)}, \quad \text{(B.5)}$$

where the terms $P(\tilde{\theta} = \theta | H, \tilde{v}^e = 0)$ and $E[\tilde{v} | H, \tilde{v}^e = 0, \tilde{\theta} = \theta]$ are given, in terms of the state variables, in Tables B.3 and B.4.

Because valuations (i.e., the entries in Table 2.2 as well as in Table B.4) are all in the open interval $(0, 1)$, the posteriors are also in the interval $(0, 1)$. This ends the constructive proof of the Markovian equilibrium.

Proof of Proposition 3. The proof proceeds in two steps. In step one, we show that the limit must be either zero or one, and in step two we show that the limit must be $\tilde{v}$.

Step One: The martingale $p^m_n = E[\tilde{v} | H_n]$ is bounded and hence converges almost surely. We fix a realization of the history of trade, H_∞, along which the martingale converges and we denote its limit by l^m.

Assume, by means of contradiction, that l^m is in the open interval $(0,1)$. Thus, also at the limit, traders of type 3 (reps. type 4) have valuations that are strictly higher (resp. lower) than the market makers' valuations (see Table 2.2). Hence, at the limit the bid-ask spread is strictly positive. Denote the spread, at the limit, by ϵ. Because of the convergence along H_∞, there is an N such that for all $n > N$, $|p_n^m - p_{n+1}^m| < \epsilon/2$. This contradicts the facts that trading occurs after N with positive probability (for example, when noise traders trade), and then p_{n+1} is either the ask or the bid. We conclude that the limit of p^m must be either one or zero.

Step Two: To see that the limit of p_n^m is almost surely $\tilde{v}$, let $\mathcal{H}_0$ be the set of all histories for which $\lim p_n^m = 0$. In particular, $P(\tilde{v} = 1|H \in \mathcal{H}_0) = 0$. Because of the symmetry of the model, $P(H \in \mathcal{H}_0) = 1/2$ and therefore:

$$P(H \in \mathcal{H}_0|\tilde{v} = 1) = P(\tilde{v} = 1|H \in \mathcal{H}_0) = 0.$$

Similarly, we show that conditional on $\tilde{v} = 0$, $\lim p_n^m = 0$.

Proof of Proposition 4. The initial price impact is the initial ask price minus $p_0^m = 1/2$ (the case is symmetric for the bid price). Thus, the claim is true if the initial ask price in the influence model is greater than the ask price in the benchmark model. The ask price in the influence model is either $E[\tilde{v}|\tilde{\theta}_1 \in \{1,3,7\}]$ or $E[\tilde{v}|\tilde{\theta}_1 \in \{1,7\}]$.[1] Either way, the ask price in the influence model is greater than $E[\tilde{v}|\tilde{\theta}_1 \in \{3,7\}]$, which is the ask price in the benchmark model, because $\pi^e > \pi$.

Proof of Proposition 5. Consider the influence model. Note that since our claim is that $\tilde{N} \geq 1$, we only need to consider the first period of trading. We conjecture and then verify that the equilibrium has the properties stated in the proposition. If the potential buyers are types 1, 2, or 7, then the ask price in the first period of trading is:

$$ask = \frac{\sum_{\theta\in\{1,2,7\}} E[\tilde{v}|\tilde{\theta} = \theta]P(\tilde{\theta} = \theta)}{\sum_{\theta\in\{1,2,7\}} P(\tilde{\theta} = \theta)} = \frac{\frac{1}{2}\frac{1-\mu}{2} + \frac{1}{2}\mu\alpha\pi^e}{\frac{1-\mu}{2} + \frac{1}{2}\mu\alpha},$$

which is strictly smaller than π^e and strictly greater than 1/2. For π sufficiently close to 1/2, $E[\tilde{v}|\theta = 2]$ is sufficiently close to π^e and hence type 2 is a buyer even though her personal view is negative. Similarly, $E[\tilde{v}|\theta = 3]$ in this case is sufficiently close to 1/2, which is smaller than the ask price and hence type 3 abstains even though her personal view is positive. In the same manner we examine the bid price and see that type 5 sells while type

[1] In Proposition 5 we show that if π^p is sufficiently close to 1/2, then the latter is the ask price.

4 abstains. Because personal views are not "expressed" during the initial stochastic window $[1, \tilde{N}]$, p_n^h and p_n^l remain constant. As for (D) and (E), we have shown numerically that they hold. An analytic proof will be provided in the next draft of the paper.

Now consider the benchmark model. Influenced investors are not present, and the model becomes a standard Glosten and Milgrom (1985) model (as long as π is strictly greater than 1/2) in which the knowledgeable investors play the role of informed traders.

Table B.1: Events and Conditional Probabilities: This table lists the four events that can be realized before trade commences.

E	Equivalently	$P(E\|H)$	$P(E\|H,\tilde{v}^e=0)$	$P(E\|H,\tilde{v}^e=1)$
$\{v=0,\epsilon^e=0\}$	$\{v=0,v^e=0\}$	$(1-p^l)(p^h-p^m)/(p^h-p^l)$	$1-p^l$	0
$\{v=0,\epsilon^e=1\}$	$\{v=0,v^e=1\}$	$(1-p^h)(p^m-p^l)/(p^h-p^l)$	0	$1-p^h$
$\{v=1,\epsilon^e=0\}$	$\{v=1,v^e=1\}$	$p^h(p^m-p^l)/(p^h-p^l)$	0	p^h
$\{v=1,\epsilon^e=1\}$	$\{v=1,v^e=0\}$	$p^l(p^h-p^m)/(p^h-p^l)$	p^l	0

Table B.2: Arrival Probabilities: This table shows the probability, conditional on the history of trading, that each type arrives in the market.

$\tilde{\theta}$	$Prob(\tilde{\theta}\|H)$
1	$\mu\alpha\left(\pi p^h+(1-\pi)(1-p^h)\right)\times(p^m-p^l)/(p^h-p^l)$
2	$\mu\alpha\left((1-\pi)p^h+\pi(1-p^h)\right)\times(p^m-p^l)/(p^h-p^l)$
3	$\mu(1-\alpha)\left(\pi p^m+(1-\pi)(1-p^m)\right)$
4	$\mu(1-\alpha)\left((1-\pi)p^m+\pi(1-p^m)\right)$
5	$\mu\alpha\left(\pi p^l+(1-\pi)(1-p^l)\right)\times(p^h-p^m)/(p^h-p^l)$
6	$\mu\alpha\left((1-\pi)p^l+\pi(1-p^l)\right)\times(p^h-p^m)/(p^h-p^l)$
7	$\frac{1}{2}(1-\mu)$
8	$\frac{1}{2}(1-\mu)$

Table B.3: Arrival Probabilities Conditional on the Published View: This table lists the types of traders together with their arrival probabilities, conditional on the history and the published view.

θ	$Prob(\tilde{\theta} = \theta \mid H, \tilde{v}^e = 0)$	$Prob(\tilde{\theta} = \theta \mid H, \tilde{v}^e = 1)$
1	0	$\mu\alpha\left(\pi p^h + (1-\pi)(1-p^h)\right)$
2	0	$\mu\alpha\left((1-\pi)p^h + \pi(1-p^h)\right)$
3	$\mu(1-\alpha)\left(\pi p^l + (1-\pi)(1-p^l)\right)$	$\mu(1-\alpha)\left(\pi p^h + (1-\pi)(1-p^h)\right)$
4	$\mu(1-\alpha)\left((1-\pi)p^l + \pi(1-p^l)\right)$	$\mu(1-\alpha)\left((1-\pi)p^h + \pi(1-p^h)\right)$
5	$\mu\alpha\left(\pi p^l + (1-\pi)(1-p^l)\right)$	0
6	$\mu\alpha\left((1-\pi)p^l + \pi(1-p^l)\right)$	0
7	$\frac{1}{2}(1-\mu)$	$\frac{1}{2}(1-\mu)$
8	$\frac{1}{2}(1-\mu)$	$\frac{1}{2}(1-\mu)$

Table B.4: Valuations Conditional on the Published View: This table lists the asset valuations associated with each trader type conditional on the published view.

θ	$E(\tilde{v} \mid H, \tilde{v}^e = 0, \tilde{\theta} = \theta)$	$E(\tilde{v} \mid H, \tilde{v}^e = 1, \tilde{\theta} = \theta)$
1	NA	$\pi p^h / (\pi p^h + (1-\pi^p)(1-p^h))$
2	NA	$(1-\pi)p^h / ((1-\pi)p^h + \pi^p(1-p^h))$
3	$\pi p^l / (\pi p^l + (1-\pi)(1-p^l))$	$\pi p^h / (\pi p^h + (1-\pi)(1-p^h))$
4	$(1-\pi)p^l / ((1-\pi)p^l + \pi(1-p^l))$	$(1-\pi)p^h / ((1-\pi)p^h + \pi(1-p^h))$
5	$\pi p^l / (\pi p^l + (1-\pi)(1-p^l))$	NA
6	$(1-\pi)p^l / ((1-\pi)p^l + \pi(1-p^l))$	NA
7	p^l	p^h
8	p^l	p^h

REFERENCES

Admati, A. R., and P. Pfleiderer, 1986, "A monopolistic market for information," *Journal of Economic Theory*, 39(2), 400–438.

Amihud, Y., 2002, "Illiquidity and stock returns: Cross-section and time-series effects," *Journal of Financial Markets*, 5(1), 31–56.

Andersen, T. G., T. Bollerslev, F. X. Diebold, and H. Ebens, 2001, "The distribution of realized stock return volatility," *Journal of Financial Economics*, 61(1), 43–76.

Banerjee, A. V., 1992, "A simple model of herd behavior," *The Quarterly Journal of Economics*, 107(3), 797–817.

Bikhchandani, S., D. Hirshleifer, and I. Welch, 1992, "A theory of fads, fashion, custom, and cultural change as informational cascades," *Journal of Political Economy*, 100(5), 992–1026.

Black, F., and R. Litterman, 1992, "Global portfolio optimization," *Financial Analysts Journal*, 48(5), 28–43.

Boehmer, E., G. Saar, and L. Yu, 2005, "Lifting the veil: An analysis of Pre-trade transparency at the NYSE," *The Journal of Finance*, 60(2), 783–815.

Bossaerts, P., 1993, "Transaction prices when insiders trade portfolios," *Finance*, 14, 43–60.

Chan, C. C., and W. M. Fong, 2006, "Realized volatility and transactions," *Journal of Banking & Finance*, 30(7), 2063–2085.

Chan, K., 1992, "A further analysis of the lead-lag relationship between the cash market and stock index futures market," *Review of Financial Studies*, 5(1), 123–152.

Chen, H., G. Noronha, and V. Singal, 2004, "The price response to S&P 500 index additions and deletions: Evidence of asymmetry and a new explanation," *The Journal of Finance*, 59(4), 1901–1930.

Davies, R. J., and S. S. Kim, 2009, "Using matched samples to test for differences in trade execution costs," *Journal of Financial Markets*, 12(2), 173–202.

Denis, D. K., J. J. McConnell, A. V. Ovtchinnikov, and Y. Yu, 2003, "S&P 500 index additions and earnings expectations," *The Journal of Finance*, 58(5), 1821–1840.

Doyle, J. T., and M. J. Magilke, 2009, "The timing of earnings announcements: An examination of the strategic disclosure hypothesis," *The Accounting Review*, 84(1), 157–182.

Easley, D., and M. O'hara, 1992, "Time and the process of security price adjustment," *The Journal of Finance*, 47(2), 577–605.

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

French, K. R., G. W. Schwert, and R. F. Stambaugh, 1987, "Expected stock returns and volatility," *Journal of Financial Economics*, 19(1), 3–29.

Gennotte, G., and B. Trueman, 1996, "The strategic timing of corporate disclosures," *Review of Financial Studies*, 9(2), 665–690.

Glosten, L. R., and P. R. Milgrom, 1985, "Bid, ask and transaction prices in a specialist market with heterogeneously informed traders," *Journal of Financial Economics*, 14(1), 71–100.

Gorton, G. B., and G. G. Pennacchi, 1993, "Security baskets and index-linked securities," *Journal of Business*, 66(1), 1–27.

Hasbrouck, J., 1995, "One security, many markets: Determining the contributions to price discovery," *The Journal of Finance*, 50(4), 1175–1199.

——— , 2003, "Intraday price formation in US equity index markets," *The Journal of Finance*, 58(6), 2375–2400.

Jiang, C. X., T. Likitapiwat, and T. H. McInish, 2012, "Information content of earnings announcements: Evidence from after-hours trading," *Journal of Financial and Quantitative Analysis*, 47(06), 1303–1330.

Jones, C. M., G. Kaul, and M. L. Lipson, 1994, "Transactions, volume, and volatility," *Review of Financial Studies*, 7(4), 631–651.

Jovanovic, B., and A. J. Menkveld, 2012, "Middlemen in limit-order markets," *Working Paper, New York University.*

Kawaller, I. G., P. D. Koch, and T. W. Koch, 1987, "The temporal price relationship between S&P 500 futures and the S&P 500 index," *The Journal of Finance*, 42(5), 1309–1329.

Kyle, A. S., 1985, "Continuous auctions and insider trading," *Econometrica*, 53(6), 1315–1335.

Lo, A. W., and J. Wang, 2000, "Trading volume: Definitions, data analysis, and implications of portfolio theory," *Review of Financial Studies*, 13(2), 257–300.

Lynch, A. W., and R. R. Mendenhall, 1997, "New evidence on stock price effects associated with changes in the S&P 500 Index," *Journal of Business*, 70, 351–383.

Merton, R. C., 1987, "A simple model of capital market equilibrium with incomplete information," *The Journal of Finance*, 42(3), 483–510.

Michaely, R., A. Rubin, and A. Vedrashko, 2014, "Corporate governance and the timing of earnings announcements," *Review of Finance*, 18(6), 2003–2044.

Patell, J. M., and M. A. Wolfson, 1982, "Good news, bad news, and the intraday timing of corporate disclosures," *Accounting Review*, 57(3), 509–527.

Qin, N., and V. Singal, 2013, "Indexing and stock price efficiency," *Working Paper, Virginia Polytechnic Institute & State University.*

Schwert, G. W., 1989, "Why does stock market volatility change over time?," *The Journal of Finance*, 44(5), 1115–1153.

Shleifer, A., 1986, "Do demand curves for stocks slope down?," *The Journal of Finance*, 41(3), 579–590.

Stoll, H. R., and R. E. Whaley, 1990, "The dynamics of stock index and stock index futures returns," *Journal of Financial and Quantitative Analysis*, 25(4), 441–468.

Subrahmanyam, A., 1991, "A theory of trading in stock index futures," *Review of Financial Studies*, 4(1), 17–51.

Surowiecki, J., 2004, "The wisdom of crowds: Why the many are smarter than the few and how collective wisdom shapes business," *Little, Brown London.*

Treynor, J. L., 1987, "Market efficiency and the bean jar experiment," *Financial Analysts Journal*, 43(3), 50–53.

Welch, I., 1992, "Sequential sales, learning, and cascades," *The Journal of Finance*, 47(2), 695–732.

Yu, L., 2005, "Basket securities, price formation, and informational efficiency," *Working Paper, University of Notre Dame.*