Word of Mouth and Gradual Information Diffusion in Asset Markets

Dong Hong
Singapore Management University

Harrison Hong
Princeton University

Andrei Ungureanu
Princeton University

Abstract

We develop a model of gradual information diffusion based on word of mouth communication. When news is initially released to a small fraction of investors, the expected diffusion rate (the change in the fraction of investors with the news) is initially increasing in time and only then decreases. Assuming risk-averse investors trade when they receive the news, the serial correlation of stock returns is proportional to the diffusion rate. We confirm the predicted non-linearity of the diffusion rate. Focusing on very large stocks where liquidity is less of an issue, we decompose returns into a component based on earnings releases (public news) and a residual (private news). The serial correlation of a return due to public news with future returns declines in time. In contrast, the serial correlation of a return due to private news with future returns initially increases and only then decreases in time.

I. Introduction

In this paper, we study how word of mouth communication affects the flow of information in the context of asset markets. Scientists have long recognized its importance in influencing the adoption of new technologies, the spread of disease, and search in labor markets (see Jackson (2009) for a review) and developed models of such social processes. A key feature of many such network models is the famous S-shaped plot of the fraction of the population affected against time. The diffusion rate (or the change in the fraction of people affected) is non-linear in time: low in the near term when there are few potential senders, high in the medium term when there are more potential senders, and low in the long term when most everyone is already affected. Yet, systematic empirical studies of how this important channel of non-market interaction affects information flow in economies are still limited.

Our focus on the stock market is motivated by a growing body of work pointing to the importance of word of mouth for investment decisions. Evidence include: (1) surveys report that investors get ideas from friends (Shiller and Pound (1989)); (2) retirement plan decisions are influenced by co-workers (Duflo and Saez (2002), Madrian and Shea (2000)); (3) stock market participation depends on friends and neighbors (Hong, Kubik and Stein (2005)); (4) trades of managers and retail investors from the same city or zip code are correlated (Hong, Kubik and Stein (2006), Ivkovic and Weisnbenner (2007), Shive (2008), Kaustia and Knupfer (2009)); (5) managers' best picks are companies whose CEOs went to same college (Cohen, Malloy and Frazzini (08)). These studies suggest that the word of mouth mechanism is relevant but do not point to the equilibrium impact of this mechanism on the dynamics of information flow.[1]

Our focus is also motivated by a large body of work on price continuation in stock markets over the last two decades. There are two main sets of facts on which there is wide agreement. First, there is price continuation after an earnings

[1] An exception is Shive (2008) who tries to test for this S-shaped pattern using the holdings data of investors in Finland but her test is limited by potential confounds. Her strategy is different from our focus on stock prices.

announcement (Bernard and Thomas (19xx, 19xx)): firms with positive surprises significantly outperform firms with negative surprises over the next six months. Second, there is price continuation after recent price changes (Jegadeesh and Titman (1993)): firms with recent (within the past year) good stock price performance significantly outperform those with recent poor stock price performance over the next year. These studies have literally spawned an entire quantitative money management industry that seeks to exploit these patterns using what are known as momentum strategies in which funds buy a portfolio of recent winners stocks and short-sell a portfolio of recent loser stocks.

There is also growing evidence to suggest that the price continuation phenomenon is due to gradual information diffusion (Hong and Stein (1999)). Stocks with less analyst coverage (and perhaps less outlets to get the news out) are more apt to exhibit price continuation (Hong, Lim and Stein (2000)). There is more price continuation during periods when more firms are announcing earnings (and perhaps analyst outlets are overloaded) (Hirshleifer and Teoh). There is cross-firm or cross-industry price continuation related to customers and suppliers (perhaps because of slow diffusion across markets or clienteles) (Menzly and Ozbas (2009) and Cohen and Frazzini, 2008). Indeed, there is even evidence that such cross-market predictability exists at the macro-level (Hong, Torous, and Valkanov (2007), Hong and Yogo (2010)).

Our main idea is to look for evidence of a word-of-mouth effect in information flows by looking at the dynamics of price continuation in stock markets. We begin by proving a few results regarding the non-linearity of the gradual information process in a canonical model of word of mouth communication with i.i.d. transmission probabilities. In this model's network set-up an initial group of friends get the "news" and each period, there is some probability one runs into a friend and passes on the information---independence is assumed across friends at a point in time and over time. This model can generate an S-shaped plot of the fraction of friends with the news against time. The intuition is the familiar one. If the initial is fraction of friends with the news is small, diffusion rate (the change in the fraction of friends with the news) is low initially (since there are few initial senders). Diffusion rate picks up as more friends have information (since more potential senders in population). But rate slows down over time as every one has information.

Our contribution here is to prove results on the non-linearity of this diffusion rate. We actually cannot prove an explicit S-shaped diffusion process---only that the diffusion rate is low initially, peaks somewhere in the medium term and decays quickly to zero in the long-term. Most papers as far as we know work with the reduced form model, known as the Bass Model and do not even attempt to prove this result. There are a few papers that derive results on diffusion rates (mostly on the expected time when everyone has the news) for the structural network model. But as far as we can tell, they do not prove results on the non-linearity of the diffusion rates.

More importantly, we embed this word of mouth model in a simple model of stock trading and pricing in which risk-averse investors only trade when they have received the information from a friend. We assume they are risk-averse or equivalently, each has a small amount of wealth and has a minimal impact on stock prices. In this setting, we show that the serial correlation of stock returns is proportional to the diffusion rate of information. Intuitively, suppose that a small fraction of investors receive good news and they buy at time 1. Price adjusts partially to this buying and we see that price has gone up between time 0 and time 1. Price continues to go up at time 2 as more investors receive the news and buy. The degree to which it increases at time 2 depends on the diffusion rate. If the diffusion rate is large then price adjusts quicker since a large fraction of investors buy at time 2 and the greater is the correlation between the price change between 0 and 1 and the price change between 1 and 2. If the diffusion rate is very small and few investors buy, then there is little price change between time 1 and 2 and little serial correlation between these consecutive returns.

Since serial correlation is proportional to the diffusion rate, the term structure of serial correlation inherits the non-linear properties of the diffusion rate. This non-linearity is reflected in the following two testable implications regarding price continuation. The first prediction is that for news not widely disseminated (and hence there are few potential senders), the serial correlation of a stock return at time t is highest for a non-overlapping return occurring at the medium term compared to the near- and long-term. The second prediction is that for news widely disseminated (and hence there are already many potential senders), the serial correlation of non-overlapping consecutive price changes declines monotonically with time.

We confirm these two predictions using stock market data. We decompose stock returns into a component based on earnings releases (public news) and a residual (private news). We then follow some standard empirical methods in the large literature on cross-sectional stock return predictability. Each month, we sort stocks into portfolios of recent winners and recent losers (based on both the component due to public news and the residual). The profits associated with buying the winners portfolio and selling the losers portfolio is proportional to the serial correlation of returns due to public news and to private news. For the returns due to public news, we find that the price continuation profits decline monotonically with horizon. That is, this strategy's profits come mostly from the near term months and decline over time. In contrast, for the returns due to private news, the profits are the greatest in the medium term.

We have to be careful that our private news residual may actually still contain a public news component since we do not have data on all available public news, only some proxies for earnings and analyst forecast reversions. As such, a clean measure of our theory is to consider a difference-in-difference estimate in which we look at the difference in the profit distributions across different months for the portfolio sorted public news and the portfolio sorted on private news. There is an economically and statistically significant difference that is consistent with our predictions. We go on to rule out a number of alternative hypotheses for our findings and develop some auxiliary tests to confirm our theory.

In this paper, we work with the explicit network model in which we model individual friends as opposed to using the reduced form Bass model for a few reasons. We do so for a few reasons. First, we are able to prove some new useful results. Second, since we cannot prove explicitly the S-shaped pattern, we need to solve the model numerically to gain some intuition for the results. We provide some useful methods for solving the model. And third, as we discuss below, an extension of this model will yield trading or turnover implications that would not be possible if we did not model individual friends and their trading behavior.

More generally, our paper contributes to a growing literature on the economics of social networks or non-market interactions, including peer effects, multiple equilibria (tipping points), and information cascades. While this literature has yielded great insights on the theoretical side, the empirical side has been plagued by identification problems (the so-called Manski critique). In other words, when we see friends make the same investment decisions, are they engaging in word of mouth or did they just get the same news. Our paper resolves to some degree this problem by testing the dynamic price implications and hence of information flow in a word of mouth model.

Our paper proceeds as follows. We develop the model in Section II. In Section III, we discuss how to calculate relevant outcomes. In Section IV, we discuss the empirical work. In Section V, we discuss the extension of our model that will yield also trading implications.

II. Model

In describing this model and the results, most of the details are relegated to the Appendix. Suppose we have n friends who each have a probability p of running into another friend each period. We assume that this is i.i.d. across friends and time. Let $G = \{1,2,...,n\}$ denote the set of friends. Suppose n_0 of the friends initially get the "news". We want to calculate the distribution over the set of friends with the news at time t. Let Π to be power set of G: $\{\{1\},...,\{n\},\{1,2\},\{1,3\},...,\{1,2,3\},...\}$. An element A of G is the set of people with the news at t. Our i.i.d. assumptions imply that the elements of G correspond to the states of a Markov Chain.

Assume that the initial distribution of friends with news is uniform over the sets A with size n_0:

$$\pi_0(A) = 1 / \binom{n}{n_0} \tag{1}$$

For simplicity we denote $q = 1 - p$ (this is the probability that a friend does not run into another friend). We next compute the transition probability from a set B to a set A:

$$\Pr(B,A) = \begin{cases} \left[1 - q^b\right]^{a-b} \cdot q^{b\cdot(n-a)} & B \subset A \\ 0 & B \not\subset A \end{cases} \tag{2}$$

The transition matrix P has these probabilities as its entries. Note that the dimension of the matrix is given by $\dim(P) = |G| = 2^n - 1$. We can now compute the distribution over the set of people with news at time t by:

$$\pi_t = \pi_0 P^t \tag{3}$$

With this, we can calculate the expected number of friends with news at time t (which we denote by e_t). This involves multiplying transition matrix. The expected fraction of friend with the news at a given time is our variable of interest. It change is the diffusion rate across time: a high diffusion rate means that this expected fraction has increased a lot from one period to the next. And using Markov Chain Theory, we can compute the expected diffusion time (τ) (i.e. the time when everybody has the information). This is the variable of interest for the literature but is less interesting for us.

The basic elements of the model can be simply illustrated with the following simple example. Suppose there are three friends. So in our notation above, this corresponds to $n = 3$, $G = \{1,2,3\}$, and $n_0 = 1$. Then the power set is given by

$$\Pi = \{\{1\},\{2\},\{3\},\{1,2\},\{1,3\},\{2,3\},\{1,2,3\}\}$$

and the prior distribution is given by

$$\pi_0 = \begin{bmatrix} \frac{1}{3} & \frac{1}{3} & \frac{1}{3} & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Some simple probability calculations yield the following transition matrix:

$$P = \begin{bmatrix} q^2 & 0 & 0 & q-q^2 & q-q^2 & 0 & (1-q)^2 \\ 0 & q^2 & 0 & q-q^2 & 0 & q-q^2 & (1-q)^2 \\ 0 & 0 & q^2 & 0 & q-q^2 & q-q^2 & (1-q)^2 \\ 0 & 0 & 0 & q & 0 & 0 & 1-q \\ 0 & 0 & 0 & 0 & q & 0 & 1-q \\ 0 & 0 & 0 & 0 & 0 & q & 1-q \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

We provide some intuition for the transition probability entries in the first row. The first row of this matrix denotes the set in which friend indexed by 1 has the information this period. The first column of this matrix denotes the set in which friend 1 has the information the next period. The probability of this occurring is that he did not meet either friend 2 or friend 3 this period. Since these are independent draws, it is given by the probability of not running into a friend (q) squared.

Continuing with this example, the second column denotes the set in which friend 2 has the information next period. The probability that friend 1 has the information this period and that only friend 2 has the information next period is 0 since we assume information once obtained cannot be lost. A similar logic explains how the third entry in the first row is 0. The four entry is the situation in which only friend 2 gets the information from friend 1 but not 3. The probability that friend 3 doesn't get the information is q and the probability that friend 2 gets the information is 1-q. Since these are independent events, the probability is q(1-q). Similar calculations yield all the other entries.

We can calculate the updated probability distribution for the sets of friends with the news by multiplying the P matrix with the prior distribution vector, which yields

$$\pi_1 = \begin{bmatrix} \frac{q^2}{3} & \frac{q^2}{3} & \frac{q^2}{3} & 2\cdot\frac{q-q^2}{3} & 2\cdot\frac{q-q^2}{3} & 2\cdot\frac{q-q^2}{3} & (1-q)^2 \end{bmatrix}$$

With this updated distribution, we can calculate the expected number of friends with the news by simply multiplying the probability of each of the sets with the number of people in each set.

$$\begin{aligned} e_1 = {} & 1\cdot\frac{q^2}{3} + 1\cdot\frac{q^2}{3} + 1\cdot\frac{q^2}{3} + 2\cdot 2\frac{q-q^2}{3} + 2\cdot 2\frac{q-q^2}{3} + 2\cdot 2\frac{q-q^2}{3} \\ & + 3\cdot(1-q)^2 = 3-2q \end{aligned}$$

With this example in hand, we now embed our word of mouth model in a simple asset-pricing model due to Hong and Stein (1999). Their paper assumes that investors trade only when they get the news---they call these agents newswatchers. This assumption is akin to investors be boundedly rational and not extracting information from prices. See Hong and Stein (2009) for a elaborate discussion of this assumption. For us, we will simply use a reduced form version of their model given by

$$P_t = e_t \varepsilon \tag{4}$$

where ε is news at time 0 is normal(0,1) and $P_{-1} = 0$. Notice that for a given shock or news, price adjustment is proportional to the fraction of investors with the news. As this fraction increases over time and price adjusts, this will lead to price continuation or serial correlation in price changes. To see this, notice that price change or stock returns is given by

$$R_{t+1} = P_{t+1} - P_t = (e_{t+1} - e_t) \cdot \varepsilon = d_t \varepsilon \tag{5}$$

where

$$d_t = e_{t+1} - e_t \tag{6}$$

Notice that this is precisely the diffusion rate or the change in the fraction of investors with the news. We will be interested in proving results regarding the non-linearity of this quantity. But before we do this, we can calculate the conditional expectation of returns in any given time in the future on the initial return when the news hit. This conditional expectation or serial correlation is given by:

$$E[R_t \mid R_0] = \frac{d_t \cdot d_0 \cdot \mathrm{var}(\varepsilon)}{d_0^2 \cdot \mathrm{var}(\varepsilon)} R_0 = \frac{d_t}{d_0} R_0 \tag{7}$$

To fix things, consider the Hong-Stein (99) gradual information diffusion model of momentum. They assume that $1/n$ investors each period get "news" (private or public). This leads to a constant diffusion rate. Notice that the initial return is 1/n given a initial shock equal to 1. Then one can calculate simply that

$$E[R_{t+k} \mid R_t] = 1/n \qquad (8)$$

for $k < n$. In other words, for fixed n, momentum profits roughly evenly distributed across k for $k < n$. The total momentum profits $(n-1/n)$ and diffusion rate $1/n$ perfectly negatively correlated.

Our model delivers non-linear diffusion rates and hence non-linearity in the serial correlations over time. To see this, we prove three theorems (all details are in the Appendix).

Theorem 1: For certain set of values for p and small enough n_0 : $d_1 > d_0$ (i.e. serial correlation of returns at medium-horizon higher than at short-horizon)

This proof relies on us calculating in closed form solutions for e_1 and e_2. The intuition is simply that as more friends get news, diffusion rate is higher.

Theorem 2 Keeping the same assumptions as in Theorem 1 and for t big enough: $d_s < d_1$ and $d_s < d_0$ for all $s > t$ (i.e. medium horizon serial correlation higher than short- and long- horizons)

Here, we use inequalities to prove that the expected number of people converges to n as t is high. The intuition is that when everyone has information, the diffusion rate is low.

Theorems 1 and 2 basically yield the non-linear diffusion rate result, which we test. Note that this result relies on a small initial group of investors having the news and that the transmission probability is small.

What these results also suggest and we make clear in Theorem 3 is that when the number of people with the news exceeds a critical threshold of n/2, then the diffusion rate declines in time.

Theorem 3: For certain sets of values for p and for $n_0 > n/2$: $d_{t+1} < d_t$ for all t (i.e. serial correlation declining over time)

Here the recursion formula for the expected number of people allows us to prove this theorem. The intuition is that if a high initial number of people get the news, similar to everyone having information in the previous setting.

III. Computation Method and Calculations

Since we cannot prove derive closed form solutions for the fraction of people with news, it is helpful to calculate the model for various parameters to confirm our Theorems and develop a feel for comparative statics. This turns out to be non-trivial for large networks. The reason is that $\dim(P) = |G| = 2^n - 1$ and hence computation becomes very difficult. To deal with this issue, we make an observation that the transition probability depends only on cardinality of the set. This will allow us to reduce the dimensionality of the problem.

First, group all sets of the same cardinality together. Define the analog of Π to be

$$\Pi^* = \{S_a : a = 1, \ldots, n\} \qquad (9)$$

where

$$S_a = \{A : |A| = a\} . \qquad (10)$$

Then define the new initial distribution over Π^* as

$$\pi_0^*(S_{|A|}) = \pi_0(A) \cdot |S_{|A|}| \qquad (11)$$

The transition probability then becomes:

$$Pr(S_b, S_a) = \binom{n-b}{a-b} Pr(B, A) \tag{12}$$

where the new transition matrix P^* has these as its elements.

Going back to our example, we can show that we obtain identical results for the expected fraction of investors with the news at any given time are the same but the setup is different. Using the method that we just outlined for the example with three friends, let $S_1 = \{\{1\},\{2\},\{3\}\}$, $S_2 = \{\{1,2\},\{1,3\},\{2,3\}\}$, and $S_3 = \{\{1,2,3\}\}$.

Then the power set is defined as

$$\Pi^* = \{S_1, S_2, S_3\}$$

and the prior distribution is given by

$$\pi_0^* = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix}$$

The new transition matrix is then given by

$$P^* = \begin{bmatrix} q^2 & 2q-2q^2 & (1-q)^2 \\ 0 & q & 1-q \\ 0 & 0 & 1 \end{bmatrix}$$

With this transition matrix and the prior distribution, we can calculate the probability distribution at time 1 as

$$\pi_1^* = \begin{bmatrix} q^2 & 2q-2q^2 & (1-q)^2 \end{bmatrix}$$

and the expected fraction of investors as

$$e_1 = 1 \cdot q^2 + 2 \cdot \left(2q - 2q^2\right) + 3 \cdot \left(1-q\right)^2 = 3 - 2q$$

Notice that the result is the same as before.

Using this method, we now calculate the expected fraction of people and the diffusion rates in Figures 1-4 to get some feel for the key predictions of our model. For these calculations we set $n = 100$ friends. We consider two scenarios. Suppose $n_0 = 1$, so that only one person initially has the news. We call this the private news scenario since few initial friends get the news. The second scenario we consider is $n_0 = 50$ (in other words, half the population has the news). We this case the public news scenario. In Figure 1, we consider the private news case. We plot e on t (for various values of p --- the transmission probability --- ranging from a high of 0.01 to a low of 0.0005. Notice that the S-shape is most apparent for intermediate values of p. For p very small, the fraction of people with the news rises very slowly. Even after 120 periods, a large fraction of the population still does not have the news. For p very large, the fraction of investors with the news increases very quickly and the S-shape gets compressed into a short time interval since the entire population gets the news very quickly when p is large enough. It is as if the S gets compressed. For intermediate values, we see the nice S-shape.

In Figure 2, we plot the diffusion rate d_t over time. Notice that we see the non-linear diffusion rate as we discussed earlier. The diffusion rate peaks somewhere in the middle periods for all cases. It is a bit hard to see the case with the smallest p since the diffusion rate peaks at the very end of the 120 periods. Figure 2 makes clear that the S-shaped pattern in levels that we see in Figure 1 corresponds to a non-linear hump shaped pattern in the diffusion rate.

In Figures 3 and 4, we plot the fraction of people with the news and the diffusion rate against time for the case second scenario of public news, respectively. For each of the transmission probabilities, we do not see and S-shape any longer in Figure 3---now we only see that the fraction increases most dramatically in the beginning and declines as everyone eventually has the news. Indeed, in Figure 4, we see that diffusion rate is highest in nearest term and monotonically declines in time, consistent with what is observed in Figure 3.

In sum, these plots provide numerical calculations for the results in Theorems 1-3. We will next look to the stock market data on price continuation to see if indeed such diffusion processed exist.

IV. Empirical Work

We test the predictions of the model by re-examining the well-known price continuation patterns associated with past price changes (momentum of Jegadeesh and Titman) and earnings announcements (post-earnings announcement drift of Bernard and Thomas). The key facts about these two phenomena are well known: at a given point in time, a portfolio of recent winners (past 1 to 12 months) or good news firms (whether measured with earnings surprises or analysts forecast revisions) outperforms a portfolio of recent losers or bad news firms over the next year by about 10%. There is no price continuation pattern after a year. What is not known is how this 10% is distributed over the year. Are the profits uniformly accumulated for each month up to a year as in the theory of Hong and Stein (1999)? Or as our theory would suggest, the patterns should differ depending on how widely released is the news: monotonically declining for more widely disseminated news like earnings announcements and hump-shaped for less widely disseminated news?

We test our predictions using a sample of all domestic common stocks listed on NYSE, AMEX or NASDAQ from January 1976 to December 2007, excluding real estate investment trusts (REITs), American Depository Receipts (ADRs), stocks with market capitalization in the bottom quantile using NYSE-AMEX breakpoints, and stocks priced below $5. We obtain daily and monthly stock returns and market capitalization data from Chicago Research in Securities Prices (CRSP), the actual earnings announcement dates and earnings from Compustat, and analyst earnings forecasts from the Institutional Brokers' Estimate System (IBES) summary file.

We utilize three pieces of public earnings news in the empirical implementation, which we obtain from the literature on price continuation following earnings announcements. The first is the cumulative market-adjusted stock returns from day -2 to day +1 around each earnings announcement date (which we denote by ABR). The idea is that the stock price reaction on an earnings date is a pure measure of the market's surprise at the release. If the market is not surprised at the firm's release, then there is no price change. The second is standardized unexpected earnings or SUE, where SUE is the earnings this quarter minus the earnings from four quarters ago divided by the standard deviation of this same unexpected earnings measure in each of the previous eight quarters (not including the current one). This earnings surprise measure adjusts for well-known seasonal patterns in corporate earnings. The downside of this measure is that one requires a total of 12 quarters to calculate this measure. The third is analyst forecast revision in month t defined as $(F_t - F_{t-1})/P_t$, where F_t is forecast consensus for fiscal year end (FY01) in month t and P_t is the price of the stock. We denote the cumulative revision in the past J months as REV(J), which is the sum of the monthly forecast revisions from month t-J+1 to month t. ABR and SUE are measures of actual announced earnings, while REV is a measure of forecasted future earnings.

At the end of each month, we want to decompose a firm's price change into a part that is explained by public news (which we measure with earnings releases or analyst revisions) and a residual component, which is not explainable by public news, which we will call private news. Here, we want to note that we use the term public versus private in that public news is know to relatively more people, not necessarily to all. There is a lot of anecdotal evidence cited in the introduction that lots of public news is not received by all market participants at the time of release. In this set-up, we view price changes or returns as being driven by a mix of public and private news. This is a simplification since some of these price changes might be due to liquidity shocks. As it turns out, we require fairly stringent data requirements to be able to perform this decomposition that only fairly large stocks are able to satisfy. We will deal with how the potential of liquidity shocks influencing price changes affects our interpretations below.

More precisely, to perform this decomposition, we run a cross-sectional regression of past J-month stock return on earnings news and forecast revisions to decompose this price change (which we call *totInfo*) into the portion explainable by *public information (pubInfo)* and a residual which is not explainable or the *private information (privInfo)* component. For example, if we use past J = 6 months returns, then we run the following cross-sectional regression:

$$Ret(J)_{i,t} = a + b_1{*}ABR(1)_{i,t} + \ldots + b_{J/3}{*}ABR(J/3)_{i,t} + c{*}REV(J)_{i,t} + e_{i,t} \qquad (13)$$

where Ret(J) is the J-month return, ABR(1)…ABR(J/3) are the ABR's within the J-month period, and REV(J) is the cumulative analyst forecast revision in the same period. Note that for J=1,2, this regression will only have one ABR. Alternatively, the ABR's can be replaced with SUE's. We obtain qualitatively identical results using SUE. The predicted value from this regression is *pubInfo* and the residual is *privInfo*.

Each month we construct information portfolios (P1 to P5) after sorting stocks into quintiles based on totInfo, pubInfo and privInfo respectively. Table 1 summarizes the firm characteristics for portfolios (P1 (recent losers or bad news), P3, and P5 (recent winners or good news)) for formation periods of J = 1, 2, 3, 6, 9, and 12. For brevity, we do not report results for P2 and P4 since they do not factor into calculations. Take J = 3 for example. Lets first consider the total information portfolios. The first column reports the number of stocks for each portfolio. There are 408 in each. The second column reports the formation period returns for each of the portfolios. We will focus on discussion on P1 and P5 portfolios since they are the ones with which we calculate the profits from momentum and post-earnings drift strategies or equivalently, the price continuation patterns. The spread between P1 (-20.79%) and P5 (33.96%) is 55%. The third column reports the return predicted by public news from the above decomposition regression for each of the portfolios. The spread between P1 (0.69) and P5 (7.78) is about 7%. The third column reports the residuals from the regression for each of the portfolios. The spread between P1 (-21.48%) and P5 (26.118%) is 48%. Roughly, we can think of the 55% spread in returns as being decomposed along the lines of 7% being due to earnings information and the remaining 48% as unexplained.

Of course, we have to caveat that we do not have all sources of public information and our private news measure is necessarily contaminated by public news. This observation will be critical later on as we will need to look at a difference-in-difference estimate between the reactions of price to public versus price news to test our hypotheses.

Continuing with the summary statistic table for total information portfolios for J=3 month returns, we also report for completeness the characteristics of the earnings data (SUE, ABR and REV) for each of the portfolios and also the LogSize of the company and the number of analyst estimates. Notice that P1 has on average poorer SUE, ABR and REV statistics compared to P3 or P5. Also notice that companies in each of the portfolios are very large due to our data restrictions. For J = 3, on average in each month 2041 stocks are included in our portfolios, 5264 stocks are in the crsp universe. So 61% of the stocks are excluded. Median market cap of the 2041 stocks is 267 million on average, which places them at the 76 percentile of the whole CRSP universe in terms of market cap. The mean market capitalization of the 2041 stocks is 1662 mil, which is the 91 percentile of the whole CRSP universe in terms of market cap. Finally, not surprisingly, these companies have on average around 7 analyst estimates each month which is much larger than the typical stock which has little to no coverage.

Notice a few other key things regarding this summary statistics table. The first is that for J=1 and 2, we have fewer stocks because to do our decomposition, we need earnings information in that month or in those 2 months and this is much less likely to occur for all stocks than within a quarter. In other words, firms are reporting their numbers at different months within a quarter and hence in any given month, we are picking up only one-third of the firms. Hence, J=3 is in some sense our preferred focus of stocks since we have a larger sample to reduce measurement error. So it should not come as a surprise that the J=1 and 2 results might be a bit noisier.

Also notice that we group together J=6, 9, 12 into a separate group of longer horizon past returns. The logic of our model is most easily applied to short horizon returns like one month or a quarter because one can plausibly think that these sorts of price changes may reflect private news not noticed by many investors. In contrast, stocks that have done extremely well or poorly for a longer horizon like 6 to 12 months, one might worry that we are now actually capturing the end of a gradual diffusion process rather than the beginning. In this instance, it is likely that perhaps many participants already have the information --- so here the diffusion rate even for what we call private news will be high initially and monotonically declining. This is a testable prediction we examine below with the J=6, 9, 12 decompositions.

Table 2 reports the return drift to past information for different formation period J=1,2,3. Panel A, B, C reports the drift to public information, private information and total information portfolios, respectively. After forming the information portfolios, we skip one month, then hold them for twelve months. The reported numbers are the average monthly return difference between P1 and P5 (the drift) in each of the twelve months and the corresponding t-statistics in italics. We can think of the return each month as being proportional to the diffusion rate of our model.

We begin by looking at the drift patterns for the public information portfolios. Notice that for J=1,2, and 3, for the return drift to public information, the drift is always the strongest in the first couple of months and then it weakens gradually over the 12-month period. This is consistent with the prediction of the word-of-mouth model to the extent that we view public news as being received by a large fraction of investors. (In our Theorem 3, the critical fraction if half of the investors.) According to our theory, the diffusion rate should be highly initially and gradually declines. This is roughly what we see. Now consider the return drift patterns to private information. Here the monthly returns are actually highest somewhere in months 5-8. They are lower in the first few months and in the last few months. Again, this is consistent with our model since for news that is not widely released, we expect a non-linear hump shaped diffusion rate or drift pattern in the data. The final panel reports the results for the total information sort. We still see a bit of a hump shaped pattern but it is not as pronounced as for the private information portfolios, which is very much to our point that the total information portfolios is a mix of reactions to public and private news and these reaction patterns are very different. But when combined, the total price reaction patterns seem more constant over time than is reality.

Table 2 Panel B presents the formal tests on the downward sloping drift to public information and the hump-shape drift to private information. As we argued earlier, it is likely the case that the private information portfolio is still contaminated with public information. One way to purge out these effects from the point of view of inference is to consider the difference in the returns each month for the public and the private information portfolios. One can think of this as a difference-in-difference estimate in which we use the shape of the public information portfolio as a control control group of what the price reaction to a pure public news shock would be. We report the results for J=1,2, and 3. To economize on space, we only report a few key statistics for the public information portfolio and the private information portfolio: the month 6 (M6) return minus the month 1 (M1) return, the month 12 minus the month 6 return, and the month 12 minus the month 1 return. We then report the difference in these differences for the private versus public portfolio. We also report the month with the largest return (which we denote by Peak) and take the difference with M1 and then look at the difference between M12 and Peak.

Starting at J=1, we see that M1 > M6 > M12 --- in other words, there is a monotonically declining return pattern for public information. Also, we report the difference in Peak minus M1 and M12 minus Peak and draw the same conclusion. In contrast, for private information, we see that M6 > M1 and M6 > M12 while M1 is basically equal to M12. More importantly, we see that Peak > M1 (a difference of 71 basis points a month with a t-statistic of 2) and Peak > M12 (with a difference of 62 basis points a month with a t-statistic of 2.71). In other words, there is a hump shaped return pattern to the private information portfolios. These differences are all the more stark when we consider the difference-in-difference estimate. Relative to what a pure public shock reaction would have been, M6 > M1 by 87 basis points with a t-statistic of 3.26 and M6 > M12 by 96 basis points with a t-statistic of 3.4. A Similar conclusion applies for Peak – M1, though the effect is smaller for Peak – M12.

Looking at J=2 and J=3, we come to pretty similar conclusions. If anything the effects are much stronger and nicer for J=2 and 3 then when compared to J=1. In part, this is because we have far fewer stocks when we look at J=1 results due fewer firms reporting earnings in any given month. We view the results in Table 2 as providing some support for our model.

In Table 3, we worry about the following alternative hypothesis. Suppose that the residual component of returns we are calling private news isn't just driven by private news but may also be capturing some liquidity component. A liquidity effect typically gives rise of reversals in the short horizon. A poor performance this month reverses itself the next month. Then what might be happening is that the returns for the private information portfolio may have low returns the first few months because of this reversal effect playing itself out. Our prior is that this effect is implausible for out sample since we are essentially looking at the largest stocks in the CRSP universe due to our data restrictions. Nonetheless, in this table, we report the results contained Panel B of Table 2 but cut by firm size. That is, we divide our sample into three groups, which we label small, medium and large firms and re-run all our analyses from Table 2 and see how the diffusion patterns vary across these sub-groups. The idea here is that we worry that liquidity effects are only a concern for smaller stocks. So if we do not see striking differences across the sample, then we can be assured that our results are not due to liquidity reversals in the short horizon.

Panel A reports the cuts for J=1. Results are noisier here not surprisingly but we can discern similar patterns across all three groups. Looking at public information portfolios, we see the monotonically declining pattern across all three cuts. Looking at private information portfolios, we also see that the middle months yield higher returns than the earlier months though the results look less stable when we compare the middle or peak months to the M12 for small and medium size stocks. Things look more stable and in the direction of our hypothesis when we look at large stocks.

This is very re-assuring since it is telling us that our effects are unlikely to be due to liquidity issues. In Panel B, we report comparable results for J=2 and in Panel C the results for J=3. J=2 results also look supportive in that there doesn't seem to be much variation across size cuts. The J=3 results in Panel C look less stable but are also generally in the same direction as our hypothesis.

Finally, in Table 4, we examine the results from Table 2 but now for J=6, 9, 12 month returns. For J=6 month results, it is surprising that they look similar to those of J=1,2, and 3. In other words, even for returns up to 6 months, it appears that there is still a non-linear hump shaped pattern in the diffusion rate. But for J=9 and 12, as we suspected, there is not discernible difference between public and private news reactions. The diffusion rates are all monotonically declining in time.

V. Conclusion

In this paper, we take a first step in identifying word-of-mouth effects for information flow in markets. Building on a canonical model of word of mouth with connected friends who have i.i.d. transmission probabilities, we derive some predictions regarding the non-linearity of price drift and hence of information diffusion to public versus private news. We tests these predictions in US stock market data and find some support for these predictions. Specifically, we find that conditioned on price moves due to public news, the diffusion rate of information is highest in the near term and declines gradually over time. In contrast, for price moves due to private news, the diffusion rate is low initially and peaks somewhere in the medium term of 5 to 9 months and then is low after.

We now suggest some avenues for further research. The key parameters for understanding information flow in this model is the size of networks n, how tight the network is (or the transmission probability p) and the initial fraction of people with the news. Further empirical work to measure these key parameters can yield potentially useful insights. In particular, earlier work by Hong, Lim and Stein (2000) showing that there is more momentum in stocks with greater analyst coverage (which can be interpreted both as a measure of the tightness of networks or of the initial fraction of people with the news) is example in this direction. Our own work has focused on very large stocks and paid more attention to the nature of initial news releases in our tests. But these two strands can perhaps be combined to yield additional insights.

Moreover, our model can be extended to allow for implications on trading activity. For instance, we can introduce a group of passive index investors into the model to trade with our friends. These passive index investors will serve as market makers who treat the trades of the friends in our model as uninformative. Then trading activity (buys or sells) is higher when the diffusion rate is higher. This means the following implications: we should see monotonically declining trading activity patterns with time after public news but hump-shaped trading activity patterns with private news. We are working on these extensions.

Appendix

A. Calculating the transition probability matrix

For two sets A and B in Π, we denote the transition probability from a set B to a set A at time t by $Pr(B,A,t)$. Note that $Pr(B,A,t)=0$ if $B \not\subset A$. We will now calculate the transition probabilities for where $B \subset A$. Unless otherwise specified we will use $a=|A|$ $b=|B|$ and $q=1-p$. Let us denote the probability that information travels from a set X to a set Y at time t by $P(X \to Y,t)$. We also denote by $P(X \not\to Y,t)$ the probability that information does not reach any person in set Y at time t, given that the people in set X have the information at time t-1. Note that X and Y do not have to be states of the Markov Chain. By the i.i.d. word-of-mouth assumption, we have

$$\Pr(B,A,t)=\prod_{i\in A-B} P(B \to \{i\},t)\cdot \prod_{j\notin A} P(B \not\to \{j\},t) \Rightarrow$$

$$\Pr(B,A,t)=\prod_{i\in A-B} \left[1-P(B \not\to \{i\},t)\right]\cdot \prod_{j\notin A} P(B \not\to \{j\},t) \Rightarrow$$

$$\Pr(B,A,t)=\prod_{i\in A-B} \left[1-\prod_{k\in B} P(\{k\} \not\to \{i\},t)\right]\cdot \prod_{j\notin A}\prod_{k\in B} P(\{k\} \not\to \{j\},t) \Rightarrow$$

$$\Pr(B,A)=\left[1-(1-p)^{|B|}\right]^{|A-B|}\cdot (1-p)^{|B|\cdot(n-|A|)} \Rightarrow$$

$$\Pr(B,A)=\left[1-q^{b}\right]^{a-b}\cdot q^{b\cdot(n-a)}$$

These transition probabilities form the elements of the transition matrix, denoted by P, of dimension $|\Pi|$. Let π_t be the probability distribution over the power set at time t. We know from the theory of Markov Chain that $\pi_t=\pi_0 \mathrm{P}^t$. One of the things we are interested in is that the expected proportion of people who have the information at time t, denoted by e_t, which is given by $e_t=\sum_{X\in\Pi}\pi_t(X)\cdot|X|$.

B. Computationally efficient solution method

A problem which arises in computing solutions to our model is that the dimension of the transition matrix is exponential in the number of friends ($\dim(\mathrm{P})=|\Pi|=2^n-1$), which makes computation difficult for a large n. As a solution, we will consider a new Markov Chain which takes advantage of the fact that we are predominantly interested in keeping track of the number of people having the information at time t and not in the exact set of people. We define the state space for this new Markov Chain to be $\Pi^* = \left\{ S_a \mid a \in \overline{1,n} \right\}$, where $S_a = \left\{ A \mid A \in \Pi; \mid A \mid = a \right\}$. If at time t the people in the set A have the information then the corresponding state of the Markov Chain is $S_{|A|}$. Furthermore the initial distribution π_0^* over the states in Π^* can be seen as being induced by the initial distribution π_0 on Π, with $\pi_0^*(S_{|A|}) = \pi_0(A) \cdot | S_{|A|} |$.

Notice that the transition probability from the previous Markov Chain: $\Pr(B,A)$ depends only on the cardinality of A and B. The probability distribution over states at time t is $\pi \cdot \mathrm{P}^t$. Since both π and P are function of sets only through their cardinality then conditioned on being in a state S_a at time t each set $A \in S_a$ is equally likely to be the one with exactly those people who have the information at time t. We are interested in computing the new transition probability $\Pr(S_b, S_a)$.

$$\Pr(S_b,S_a) = \sum_{A \in S_a} \sum_{B \in S_b} \frac{1}{|S_b|} \Pr(B,A) = \sum_{A \in S_a} \sum_{B \in S_b; B \subset A} \frac{1}{|S_b|} \Pr(B,A) \Rightarrow$$

$$\Pr(S_b,S_a) = \sum_{|A|=a} \sum_{B \subset A; |B|=b} \frac{1}{|S_b|} \Pr(b,a) = \frac{\Pr(b,a)}{|S_b|} \cdot \sum_{|A|=a} \sum_{B \subset A; |B|=b} 1 \Rightarrow$$

$$\Pr(S_b,S_a) = \frac{\Pr(b,a)}{|S_b|} \cdot \sum_{|A|=a} \begin{pmatrix} a \\ b \end{pmatrix} = \frac{\Pr(b,a)}{|S_b|} \cdot \begin{pmatrix} n \\ a \end{pmatrix} \cdot \begin{pmatrix} a \\ b \end{pmatrix} \Rightarrow$$

$$\Pr(S_b,S_a) = \begin{pmatrix} n-b \\ a-b \end{pmatrix} \cdot \Pr(b,a) = \begin{pmatrix} n-b \\ a-b \end{pmatrix} \cdot \left[1-q^b\right]^{a-b} \cdot q^{b \cdot (n-a)}$$

Since people do not forget information once they receive it $\Pr(S_i,S_j) = 0$ for all $i > j$. The transition matrix P^* is: $p_{ij}^* = \Pr(S_i,S_j)$ if $i < j$ and $p_{ij}^* = 0$ otherwise, $\forall i,j \in \overline{1,n}$. It is worth noticing that the dimension of the transition matrix is now n, linear as opposed to exponential in the number of agents of the network. We can now write down the formula for the distribution over the states of the new Markov Chain at time t: $\pi_t^* = \pi_0^* \cdot \left(\mathrm{P}^*\right)^t$. It follwos that the expected number of people who have the information at time t, denoted by e_t^*, is: $e_t^* = \sum_{S_a \in \Pi^*} \pi_t^*(S_a) \cdot a = e_t$.

C. Results need for proofs of Theorems 1-3

We now prove some results which will be used in the proofs of Theorems 1-3. Since we are only interested in computing the expected number of friends with the information at a particular time, we will use the framework for the computational efficient Markov Chain model. Let $g = \begin{bmatrix} 1 & 2 & \dots & n \end{bmatrix}'$ and π_0^* be the initial distribution as above.

Let us define $a_m(t)$ as the expected number of people that do not have the information at time t given that m people do not have the information at time 0. Notice that , trivially for all t since when everybody has the information at time 0, everybody will have the information at any other time in the future.

Proposition 1: *For all* $m \in \{1,\dots,n-1\}$ *there exists a positive real number* u *such that the following recursion holds:*

$$a_m(t+1) = \sum_{k=0}^{m} u^{m-k} \cdot v^k \cdot \begin{pmatrix} m \\ k \end{pmatrix} \cdot a_{m-k}(t)$$

where $v = 1 - u$.

Proof. Using the definition of the Markov Chain and that of $a_m(t)$ we have:

$$\left(\mathrm{P}^*\right)^t \cdot g = n \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} - \begin{pmatrix} a_{n-1}(t) \\ \vdots \\ a_0(t) \end{pmatrix}$$

Because P^* is a transition matrix for a Markov Chain the following holds $\left(\mathrm{P}^*\right)^t \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$.

We can now try to derive a recursive formula by looking at the one step update in this matrix notation:

$$n \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} - \begin{pmatrix} a_{n-1}(t+1) \\ \vdots \\ a_0(t+1) \end{pmatrix} = \left(\mathrm{P}^*\right)^{t+1} \cdot g = \mathrm{P}^* \cdot \left(n \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} - \begin{pmatrix} a_{n-1}(t) \\ \vdots \\ a_0(t) \end{pmatrix} \right) = n \cdot \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix} - \mathrm{P}^* \cdot \begin{pmatrix} a_{n-1}(t) \\ \vdots \\ a_0(t) \end{pmatrix} \Rightarrow$$

$$\mathrm{P}^* \cdot \begin{pmatrix} a_{n-1}(t) \\ \vdots \\ a_0(t) \end{pmatrix} = \begin{pmatrix} a_{n-1}(t+1) \\ \vdots \\ a_0(t+1) \end{pmatrix} \Rightarrow \mathrm{P}_i^* \cdot \begin{pmatrix} a_{n-1}(t) \\ \vdots \\ a_0(t) \end{pmatrix} = a_{n-i}(t+1)$$

where we have denoted by P_i^* the i^{th} line of the matrix P^*.

Writing explicitly the formulas for the entries of P^* we get:

$$a_{n-i}(t+1) = \sum_{k=0}^{n-i} (q^i)^{n-i-k} \cdot (1-q^i)^k \cdot \begin{pmatrix} n-i \\ k \end{pmatrix} \cdot a_{n-i-k}(t)$$

Finally we can make the notation $u = q^i$, $v = 1-u$, $m = n-i$ and the desired recursion formula is proven:

$$a_m(t+1) = \sum_{k=0}^{m} u^{m-k} \cdot v^k \cdot \begin{pmatrix} m \\ k \end{pmatrix} \cdot a_{m-k}(t)$$

Corollary 1: *The recursion stated in the previous proposition holds for the following sequences as well:* $d_m(t) = \left[n - a_m(t+1)\right] - \left[n - a_m(t)\right]$ *and* $dd_m(t) = d_m(t+1) - d_m(t)$.

Proof. The proof is straight forward. All we need to do is to explicitate $d_m(t)$ using the recursion for $a_m(t+1)$ and $a_m(t)$. Once the first claim is proved we use the same method for the second claim.

Proposition 2: *The following close form formulas hold for* $m \in \{1,...,n-1\}$:

1. $a_m(0) = m$
2. $a_m(1) = m \cdot q^{n-m}$
3. $a_m(2) = m \cdot q^{2(n-m)} \cdot \left[q^{n-m}(1-q) + q\right]^{n-1}$

Proof. Since we know that $g = \begin{bmatrix} 1 & 2 & \dots & n \end{bmatrix}'$ we can immediately derive $a_m(0) = m$ for $m \in \{1,...,n-1\}$.

We will use the recursion formula and the fact that $a_m(0) = m$ to derive the equality for $a_m(1)$.

$$a_m(1) = \sum_{k=0}^{m} u^{m-k} \cdot v^k \cdot \begin{pmatrix} m \\ k \end{pmatrix} \cdot a_{m-k}(0) \Rightarrow a_m(1) = \sum_{k=0}^{m} u^{m-k} \cdot v^k \cdot \begin{pmatrix} m \\ k \end{pmatrix} \cdot (m-k) \Rightarrow$$

$$a_m(1) = \sum_{k=0}^{m-1} u \cdot u^{m-1-k} \cdot v^k \cdot \begin{pmatrix} m-1 \\ k \end{pmatrix} \cdot m \Rightarrow a_m(1) = u \cdot m \cdot (u+v)^{m-1} \Rightarrow$$

$$a_m(1) = m \cdot q^{n-m}$$

Let us now try to compute $a_m(2)$ in the same manner:

$$a_m(2) = \sum_{k=0}^{m} u^{m-k} \cdot v^k \cdot \begin{pmatrix} m \\ k \end{pmatrix} \cdot a_{m-k}(1) \Rightarrow a_m(2) = \sum_{k=0}^{m} u^{m-k} \cdot v^k \cdot \begin{pmatrix} m \\ k \end{pmatrix} \cdot (m-k) \cdot q^{n-m+k} \Rightarrow$$

$$a_m(2) = \sum_{k=0}^{m-1} u \cdot u^{m-1-k} \cdot (q \cdot v)^k \cdot \begin{pmatrix} m-1 \\ k \end{pmatrix} \cdot m \cdot q^{n-m} \Rightarrow a_m(2) = u \cdot m \cdot q^{n-m} \cdot (u+qv)^{m-1} \Rightarrow$$

$$a_m(2) = m \cdot q^{2(n-m)} \cdot \left[q^{n-m}(1-q) + q \right]^{m-1}$$

Remark 1: Given our assumption about the initial distribution, the expected number of people having the information at time t is increasing in t. In our notation this amounts to $d_m(t) > 0$ for all m and t, which is obviously true because $d_m(0) = m\left(1 - q^{n-m}\right) > 0$ and the recursion coefficients are positive. In the next proposition, we prove this result for a general π_0^*.

Proposition 3: *Given any initial probability distribution π_0^* over the number of people having the information at time 0, the expected number of people having the information at time t is a strictly increasing function of t.*

Proof. Let $a(t) = \begin{bmatrix} a_{n-1}(t) & \dots & a_0(t) \end{bmatrix}'$ and $d(t) = \begin{bmatrix} d_{n-1}(t) & \dots & d_0(t) \end{bmatrix}'$. All we need to prove is that $\pi_0^* \cdot d(t+1) > 0$ for all t. Using the definition of $d(t+1)$ this is equivalent to $\pi_0^* \cdot (a(t) - a(t+1)) > 0$ for all t and for all initial distributions π_0^*. We know that

$$a(t+1) = \mathrm{P}^* \cdot a(t) \Rightarrow a(t) = \left(\mathrm{P}^*\right)^t \cdot a(0) \Rightarrow a(t) - a(t+1) = \left(\mathrm{P}^*\right)^t \cdot \left[a(0) - a(1)\right].$$

Using Proposition 2 we have $a_m(0) - a_m(1) = m \cdot (1 - q^{n-m})$ which is strictly positive for all $m \in \{1, \dots, n-1\}$ and 0 for $m = 0$. On the other hand P^* is an upper triangular matrix with strictly positive entries. This readily implies that $\left(P^*\right)^t$ is also an upper triangular matrix with strictly positive entries. Simple matrix multiplication leads to the fact that the vector $a(t) - a(t+1) = \left(\mathrm{P}^*\right)^t \cdot \left[a(0) - a(1)\right]$ has strictly positive entries, except for the last one which is 0.

Since π_0^* is a probability distribution it has only nonnegative elements and since $\pi_0^*(n)=0$ $\exists i \in \{1,...,n-1\}$ such that $\pi_0^*(i)>0$. This implies that $\pi_0^* \cdot [a(t)-a(t+1)]$ is strictly positive.

D. Proof of Theorem 1

We want to show the following: given k a positive integer, $dd_m(0)>0$ for all $m>n-\bar{n}_0$ (where $\bar{n}_0 = \left[\frac{n-1}{k+1}\right]$) assuming $q^{n_0-1} < \frac{k-1}{k}$ and $q^{\bar{n}_0}+q^{2\bar{n}_0}>1$.

First, using Proposition 2, observe that

$$dd_m(0) = 2q^{n_0} - 1 - q^{2n_0}\left[q^{n_0}(1-q)+q\right]^{n-n_0-1},$$

where $n_0 = n-m$. Now we want to show that $\left[q^{n_0}(1-q)+q\right]^{n-n_0-1} < q^{n_0}$. Initially we will prove that $\left[q^{n_0}(1-q)+q\right]^{k} < q$. Let $x^k = q$, where $x>0$. Let $y=\sum_{i=0}^{k-2} x^i$. Since $q<1 \Rightarrow x<1$, we have

$$y > (k-1)x^{k-1} \Rightarrow ky > (k-1)(y+x^{k-1}) \Rightarrow \frac{1-x^{k-1}}{1-x^k} > \frac{k-1}{k}.$$

Using $q^{n_0-1} < \frac{k-1}{k}$ we have

$$\frac{1-x^{k-1}}{1-x^k} > q^{n_0-1} > x^{kn_0-1} \Rightarrow x^{kn_0}(1-x^k) < x - x^k \Rightarrow \left[q^{n_0}(1-q)+q\right]^{k} < q.$$

Observe that $n_0 < \overline{n_0} \le \frac{n-1}{k+1} \Rightarrow kn_0 < n-n_0-1$. Since $q^{n_0}(1-q)+q<1$ we derive: $\left[q^{n_0}(1-q)+q\right]^{n-n_0-1} < \left[q^{n_0}(1-q)+q\right]^{kn_0} < q^{n_0}$.

Finally we can go back to our target quantity:

$$dd_m(0) = 2q^{n_0} - 1 - q^{2n_0}\left[q^{n_0}(1-q)+q\right]^{n-n_0-1} \Rightarrow$$

$$dd_m(0) > 2q^{n_0} - 1 - q^{2n_0}q^{n_0} = (1-q^{n_0})(q^{n_0}+q^{2n_0}-1) > 0.$$

The last step follows from $q<1$, $n_0 < \overline{n_0}$ and $q^{\bar{n}_0}+q^{2\bar{n}_0}>1$.

E. Proof of Theorem 2

We want to show the following inequalities hold for any $m \in \{1,...,n-1\}$ and $t>0$:

1. $a_m(t) \le m \cdot q^{(n-m)\cdot t}$
2. $a_m(t) \ge m \cdot q^{(n-m)\cdot t} \cdot q^{(m-1)\cdot(t-1)}$

The proof goes by induction on t. From Proposition 2 we know that the two inequalities hold for $t = 1$, $t = 2$ and for all $m \in \{1,...,n-1\}$. Now suppose the inequalities hold for a given t and all $m \in \{1,...,n-1\}$ and let us prove that they hold for $t+1$ as well.

We start from the recursion formula and we will use the first inequality for each $a_{m-k}(t)$.

$$a_m(t+1) = \sum_{k=0}^{m} u^{m-k} \cdot v^k \cdot \binom{m}{k} \cdot a_{m-k}(t) \Rightarrow$$

$$a_m(t+1) \le \sum_{k=0}^{m} u^{m-k} \cdot v^k \cdot \binom{m}{k} \cdot (m-k) \cdot q^{(n-m+k)\cdot t} \Rightarrow$$

$$a_m(t+1) \le \sum_{k=0}^{m-1} u^{m-1-k} \cdot (v \cdot q^t)^k \cdot \binom{m-1}{k} \cdot u \cdot m \cdot q^{(n-m)\cdot t} \Rightarrow$$

$$a_m(t+1) \le u \cdot m \cdot q^{(n-m)\cdot t} \cdot (u + v \cdot q^t)^{m-1}$$

But $v = 1-u$ and $q < 1$ so this implies $u + v \cdot q^t < 1$. Furthermore $u = q^{n-m}$ so we get the first inequality for $t+1$.

$$a_m(t+1) \le m \cdot q^{(n-m)\cdot(t+1)}$$

The idea for the proof of the second inequality is the same: start with the recursion formula and apply the second inequality for $a_{m-k}(t)$.

$$a_m(t+1) = \sum_{k=0}^{m} u^{m-k} \cdot v^k \cdot \binom{m}{k} \cdot a_{m-k}(t) \Rightarrow$$

$$a_m(t+1) \ge \sum_{k=0}^{m} u^{m-k} \cdot v^k \cdot \binom{m}{k} \cdot (m-k) \cdot q^{(n-m+k)\cdot t} \cdot q^{(m-k-1)\cdot(t-1)} \Rightarrow$$

$$a_m(t+1) \ge \sum_{k=0}^{m-1} u^{m-1-k} \cdot (v \cdot q^t)^k \cdot \binom{m-1}{k} \cdot u \cdot m \cdot q^{(n-m)\cdot t} \cdot q^{(m-1-k)\cdot(t-1)} \Rightarrow$$

$$a_m(t+1) \ge q^{n-m} \cdot m \cdot q^{(n-m)\cdot t} \sum_{k=0}^{m-1} (u \cdot q^{t-1})^{m-1-k} \cdot (v \cdot q^t)^k \cdot \binom{m-1}{k} \Rightarrow$$

$$a_m(t+1) \ge m \cdot u \cdot q^{(n-m)\cdot t} (u \cdot q^{t-1} + v \cdot q^t)^{m-1}$$

But we know that $u = q^{n-m}$, $v = 1-u$ and $q < 1$ so it follows that:

$$a_m(t+1) \geq m \cdot q^{(n-m)\cdot(t+1)}(u \cdot q^t + v \cdot q^t)^{m-1} \Rightarrow$$

$$a_m(t+1) \geq m \cdot q^{(n-m)\cdot(t+1)} q^{(t+1-1)\cdot(m-1)}$$

This completes the induction step for the second inequality.

The inequalities derived above assure that the expected number of people that have the information at time t converges to n exponentially fast as t goes to infinity for any initial distribution π_0^* and a fixed number of agents n: $\lim_{t\to\infty} e_t^* = 1$. This assures a nonlinearity of the function e_t^* in t.

F. Proof of Theorem 3

There are two facts we need to show to prove this theorem.

Fact 1: If $dd_m(0) < 0 \quad \forall m < \overline{m}$ then $dd_m(t) < 0 \quad \forall t$ and $m < \overline{m}$.

This claim is easy to prove by induction. Assume the conclusion holds for t and let us look what happens at time $t+1$. The recursion formula says that $dd_{m_0}(t+1)$ is a linear combination of $dd_m(t)$ for m's less than m_0. Knowing that the coefficients are positive and that $m_0 < \overline{m} \Rightarrow m \leq m_0 < \overline{m}$, we can use the induction step to reach $d_{m_0}(t) > 0$ for any $m_0 < \overline{m}$.

Fact 2: *Given that* $q^n + q^{n/2} < 1$, $dd_m(0) < 0$ *for any* $m < n/2$.

Using Proposition 2 the following formula holds:

$$dd_m(0) = 2q^{n_0} - 1 - q^{2n_0}\left[q^{n_0}(1-q) + q\right]^{n-n_0-1},$$

where $n_0 = n - m$. We need to prove that $2q^{n_0} < 1 + q^{2n_0}\left[q^{n_0}(1-q) + q\right]^{n-n_0-1}$.

First let us show that $\left[q^{n_0}(1-q) + q\right]^{n-n_0-1} > \left[q^{n_0}(1-q) + q\right]^{n_0}$. This follows immediately from the following two observations: $q^{n_0}(1-q) + q < (1-q) + q = 1$ and $m < n/2 \Rightarrow n_0 = n - m > n/2 \Rightarrow n_0 > n - n_0 - 1$.

We are left to prove that $2q^{n_0} < 1+q^{2n_0}\left[q^{n_0}(1-q)+q\right]^{n_0}$. Notice that $1+q^{2n_0}\left[q^{n_0}(1-q)+q\right]^{n_0} > 1+q^{2n_0}q^{n_0} = 1+q^{3n_0}$. Since $n_0 > n/2$ and $q<1$ we have $q^{n_0}+q^{n_0/2} < q^n + q^{n/2} < 1 \Rightarrow (1-q^{n_0})(q^{n_0}+q^{n_0/2}) < 1-q^{n_0} \Rightarrow 1+q^{3n_0} > 2q^{n_0}$.
This concludes the proof of the second fact.

These two facts then imply: given that $q^n + q^{n/2} < 1$, $dd_m(t) < 0$ for any $m < n/2$ and for all t, which is a restatement of our theorem.

References

Bernard, Victor, and Jacob Thomas, 1989, Post-earnings-announcement drift: delayed price response or risk premium?, *Journal of Accounting Research* 27, 1-48.

Bernard, Victor, and Jacob Thomas, 1990, Evidence that stock prices do not fully reflect the implications of current earnings for future earnings, *Journal of Accounting and Economics*, 305-340.

Cohen, Lauren, and Andrea Frazzini, 2008, Economic links and predictable returns, *Journal of Finance* 63, 1977-2011.

Cohen, Lauren, Andrea Frazzini and Christopher Malloy, 2008, The small world of investing: board connections and mutual fund returns, *Journal of Political Economy* 116, 951-979.

Duflo, Esther, and Emmanuel Saez, 2002, Participation and investment decisions in a retirement plan: The influence of colleagues' choices, *Journal of Public Economics* 85, 121–148.

Hirshleifer , David, Sonya Lim and Siew Hong Teoh, 2007, Driven to Distraction: Extraneous Events and Underreaction to Earnings News, *Journal of Finance*, forthcoming.

Hong, Harrison, Jeffrey D. Kubik and Jeremy C. Stein, 2004, Social Interaction and Stock Market Participation, *Journal of Finance* 59, 137-163.

Hong, Harrison, Jeffrey D. Kubik and Jeremy C. Stein, 2005, Thy Neighbor's Portfolio: Word-of-Mouth Effects in the Holdings and Trades of Money Managers, *Journal of Finance* 60, 2801-2824.

Hong, Harrison, Terence Lim and Jeremy C. Stein, 2000, Bad News Travels Slowly: Size, Analyst Coverage and the Profitability of Momentum Strategies, *Journal of Finance* 55, 265-296.

Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, and overreaction in asset markets, *Journal of Finance* 54, 2143–2184.

Hong, Harrison, Walter Torous and Rossen Valkanov, 2007, Do industries Lead Stock Markets?, *Journal of Financial Economics* 83, 367-396.

Hong, Harrison and Motohiro Yogo, 2010, Digging into commodities, working paper.

Ivković, Zoran, and Scott Weisbenner, 2007, Information diffusion effects in individual investors' common stock purchases: covet thy neighbors' investment choices, *Review of Financial Studies* 20, 1327-1357.

Jackson, Matthew, 2008, Social and economics networks, *Princeton University Press*.

Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: Implications for stock market efficiency, *Journal of Finance* 48, 65–91.

Kaustia, Markku, and Samuli Knupfer, 2009, Peer performance and stock market entry, working paper.

Madrian, Brigette, and Dennis Shea, 2000, Peer effects and savings behavior in employer sponsored savings plans, University of Chicago working paper.

Menzly, Lior, and Oguzhan Ozbas, 2009, Market segmentation and cross-predictability of returns, *Journal of Finance*, forthcoming.

Shiller, Robert, and John Pound, 1989, Survey evidence on diffusion of interest and information among investors, *Journal of Economic Behavior and Organization* 12, 47–66.

Shive, Sohie, 2008, An Epidemic model of investor behavior, *Journal of Financial and Quantitative Analysis,* forthcoming.

Figure 1 : Plot of the expected number of people with the news against time

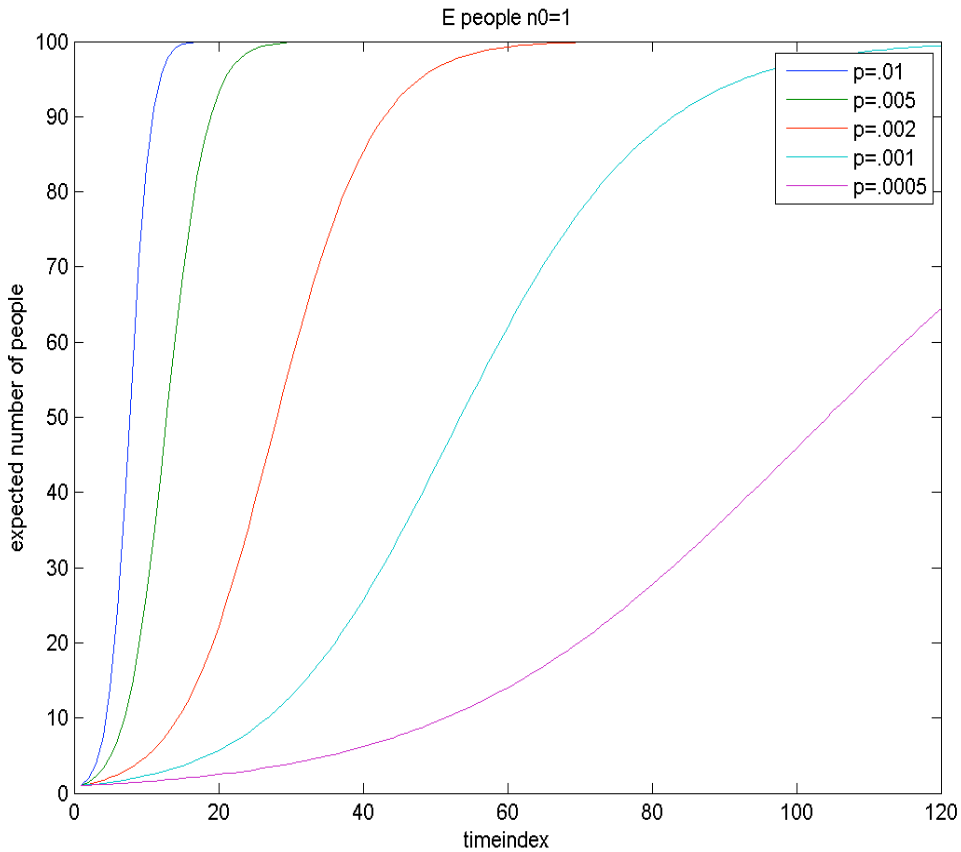

Figure 2: Plot of the expected incremental number of people with the news against time

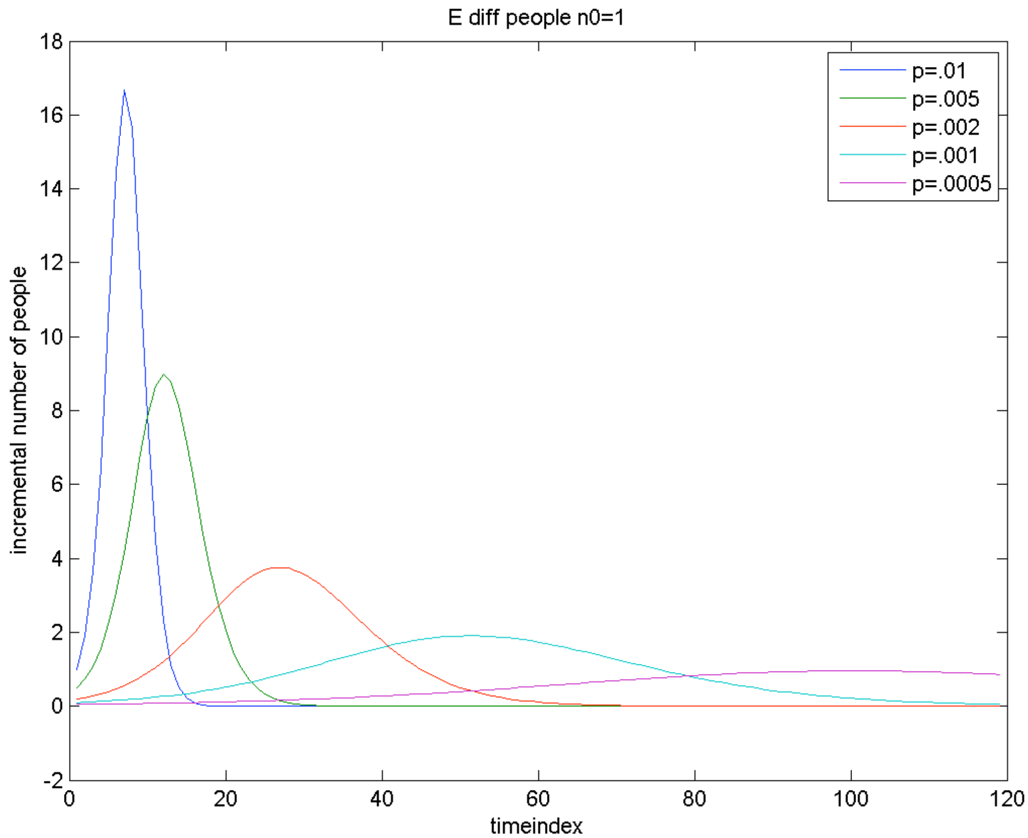

Figure 3: Plot of the expected number of people with the news against time

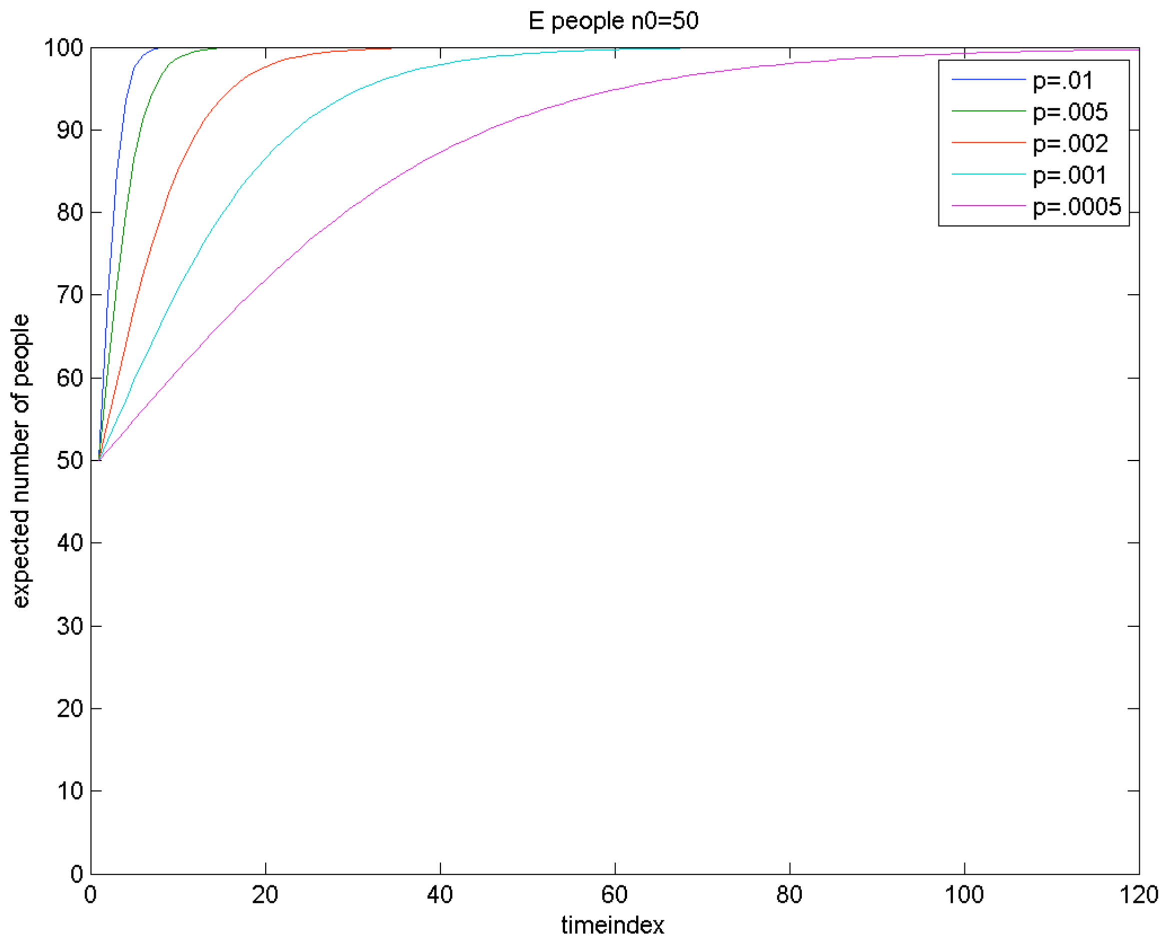

Figure 4: Plot of the expected incremental number of people with the news against time

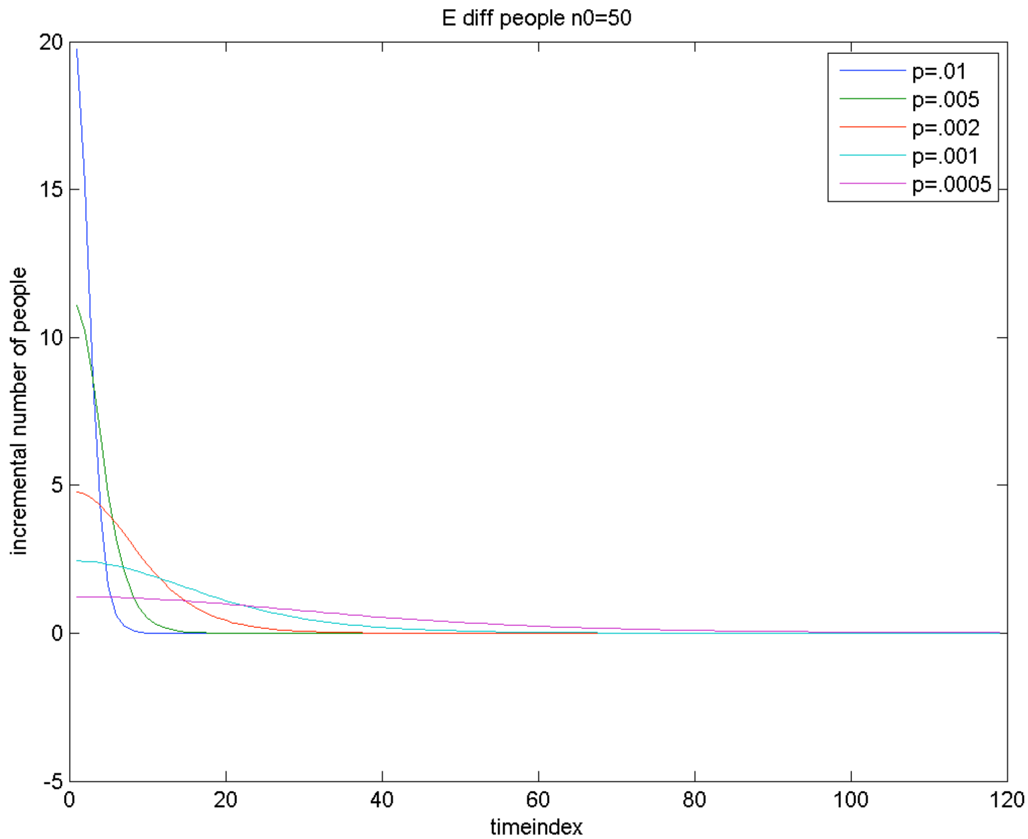

Table 1: Information Portfolio Characteristics

This table summarizes the firm characteristics for each information portfolio using various formation periods ranging from one month to twelve months. At the end of each month, we use the past J-month stock returns to proxy for total information (totInfo), which is then decomposed into public information (pubInfo, stock returns that are explained by earnings news in the past J-month) and private information (privInfo, stock returns that are orthogonal to earnings news in the past J-month) components as described in section x.xx. The stocks are then assigned to quintiles according to totInfo, pubInfo, or privInfo. The numbers reported are the time series means of the average characteristics for each portfolio each month. numStks is the average number of stocks in each portfolio. SUE is the average standardized unexpected quarterly earnings in the formation period. ABR is the average abnormal returns from day t-2 to t+1 for each earnings announcement in the formation period. REV is the average monthly revision in the analysts' FY0 forecast consensus in the formation period, deflated by the stock price. logSize is the log of the stock market capitalization (in thousands of dollars). And finally numEst is the number of analysts providing earnings forecasts for each stock. The sample period is from January 1972 to December 2007.

			numStks	totInfo	pubInfo	privInfo	SUE	ABR	REV	LogSize	numEst
J=1	Total Info	P1	157	-14.62	-2.94	-11.68	-0.48	-4.81	-1.0352	11.89	5.84
		P3	158	1.05	1.90	-0.85	0.16	0.38	-0.1904	12.52	7.13
		P5	157	20.30	6.37	13.93	0.43	5.19	-0.0976	12.25	6.00
	Public Info	P1	157	-6.64	-6.78	0.14	-0.54	-9.08	-1.3271	11.96	5.91
		P3	157	1.66	1.73	-0.08	0.16	0.17	-0.1851	12.54	7.12
		P5	157	10.75	10.68	0.07	0.48	10.02	0.0799	12.14	5.95
	Private Info	P1	157	-11.66	2.36	-14.02	-0.24	0.82	-0.4904	11.91	5.76
		P3	157	0.78	1.46	-0.68	0.13	-0.08	-0.4424	12.53	7.21
		P5	157	18.02	2.21	15.81	0.24	0.75	-0.4327	12.22	5.97
J=2	Total Info	P1	306	-18.28	-0.77	-17.51	-0.40	-2.94	-0.0124	11.99	6.40
		P3	306	1.98	3.29	-1.30	0.18	0.36	-0.0020	12.70	7.84
		P5	306	27.93	6.90	21.03	0.42	3.44	-0.0002	12.40	6.45
	Public Info	P1	306	-5.70	-5.75	0.05	-0.44	-7.11	-0.0166	12.07	6.34
		P3	306	3.07	3.13	-0.05	0.20	0.19	-0.0016	12.74	7.91
		P5	306	12.29	12.25	0.04	0.46	8.02	0.0027	12.26	6.39
	Private Info	P1	306	-15.86	3.77	-19.64	-0.20	0.73	-0.0061	12.01	6.35
		P3	306	1.73	2.83	-1.10	0.17	-0.01	-0.0027	12.71	7.88
		P5	306	26.16	3.57	22.58	0.28	0.76	-0.0049	12.38	6.42
J=3	Total Info	P1	408	-20.79	0.69	-21.48	-0.35	-2.21	-0.0191	12.05	6.76
		P3	408	2.83	4.56	-1.74	0.23	0.34	-0.0030	12.85	8.45
		P5	408	33.96	7.78	26.18	0.46	2.72	-0.0005	12.54	6.89
	Public Info	P1	408	-5.08	-4.90	-0.18	-0.41	-6.11	-0.0257	12.14	6.67
		P3	408	4.36	4.42	-0.05	0.27	0.20	-0.0021	12.88	8.54
		P5	408	13.85	13.70	0.14	0.50	6.94	0.0039	12.37	6.81
	Private Info	P1	408	-18.55	5.02	-23.57	-0.16	0.64	-0.0096	12.08	6.77
		P3	408	2.57	4.05	-1.48	0.21	0.01	-0.0037	12.85	8.46
		P5	408	32.38	4.81	27.57	0.33	0.76	-0.0075	12.52	6.88

Table 1 Cont'd

			numStks	totInfo	pubInfo	privInfo	SUE	ABR	REV	LogSize	numEst
J=6	Total Info	P1	375	-26.86	1.93	-28.79	-0.46	-1.86	-0.0401	12.01	6.98
		P3	375	5.72	9.10	-3.38	0.27	0.31	-0.0045	12.94	8.82
		P5	375	52.84	14.84	38.00	0.62	2.45	0.0034	12.69	7.27
	Public Info	P1	375	-7.05	-6.68	-0.37	-0.52	-5.11	-0.0509	12.12	6.78
		P3	375	8.71	8.99	-0.28	0.30	0.22	-0.0036	13.01	9.07
		P5	375	24.72	23.66	1.06	0.59	5.91	0.0100	12.44	7.00
	Private Info	P1	375	-22.55	10.38	-32.94	-0.19	0.95	-0.0176	12.10	7.10
		P3	375	5.17	7.92	-2.75	0.23	-0.08	-0.0078	12.92	8.78
		P5	375	49.98	9.59	40.39	0.46	0.71	-0.0113	12.66	7.23
J=9	Total Info	P1	347	-30.75	3.09	-33.84	-0.55	-1.61	-0.0596	11.99	7.14
		P3	347	8.56	13.95	-5.40	0.28	0.31	-0.0066	13.02	9.20
		P5	347	70.69	22.24	48.45	0.71	2.23	0.0061	12.81	7.57
	Public Info	P1	347	-8.31	-8.03	-0.28	-0.56	-4.28	-0.0719	12.11	6.94
		P3	347	12.97	13.78	-0.81	0.32	0.23	-0.0049	13.10	9.46
		P5	347	36.09	33.61	2.48	0.64	5.08	0.0130	12.52	7.26
	Private Info	P1	347	-24.14	16.39	-40.53	-0.21	1.11	-0.0246	12.12	7.37
		P3	347	7.62	11.81	-4.19	0.21	-0.15	-0.0118	12.98	9.10
		P5	347	66.44	14.54	51.90	0.54	0.68	-0.0153	12.76	7.51
J=12	Total Info	P1	348	-33.92	4.08	-38.00	-0.59	-1.42	-0.0790	11.95	7.14
		P3	348	11.11	18.77	-7.66	0.30	0.30	-0.0088	13.07	9.43
		P5	348	88.77	29.94	58.83	0.78	2.08	0.0083	12.90	7.82
	Public Info	P1	348	-9.65	-9.52	-0.13	-0.58	-3.73	-0.0913	12.10	6.99
		P3	348	17.03	18.68	-1.65	0.33	0.24	-0.0075	13.15	9.72
		P5	348	47.85	43.70	4.15	0.69	4.52	0.0154	12.59	7.47
	Private Info	P1	348	-24.82	22.81	-47.63	-0.21	1.22	-0.0316	12.13	7.48
		P3	348	9.83	15.69	-5.86	0.22	-0.16	-0.0161	13.02	9.28
		P5	348	82.91	19.49	63.41	0.59	0.64	-0.0184	12.84	7.72

Table 2: Return Drift to Short-term Information

This table reports the drift to short term total, public, and private information. I.e. formation period = 1, 2 and 3 months. After forming the information portfolios, we skip one month, then hold them for twelve months. In panel A, the reported numbers are the average monthly return of the hedged portfolio where we long the top information quintile and short the bottom quintile, and the corresponding t-statistics in italics. Panel B tests the shape of the drift to public information and private information. Namely, we investigate the differences of the returns of the hedged portfolio in month 1, 6, 12 and the peak month when the maximum is obtained in the drift to private information. Such differences are calculated for both the drifts to public information and to private information. Then we calculate the corresponding difference-in-difference to test the difference of the shapes of the two drifts. The differences in panel B are serially correlated and we use Newey-West t-statistics to adjust for the serial correlation.

Panel A: Return Drift to Total, Public and Private Information

Month	1	2	3	4	5	6	7	8	9	10	11	12
	Return Drift to Public Information											
J=1	1.07	1.08	0.20	0.35	0.63	0.66	0.47	0.46	0.01	-0.04	0.20	0.20
	(7.12)	*(7.19)*	*(1.36)*	*(2.15)*	*(4.12)*	*(4.44)*	*(2.99)*	*(2.90)*	*(0.08)*	*(-0.27)*	*(1.25)*	*(1.67)*
J=2	1.09	0.71	0.26	0.43	0.56	0.55	0.39	0.16	0.08	0.17	0.24	0.08
	(9.21)	*(6.51)*	*(2.15)*	*(3.84)*	*(4.82)*	*(4.61)*	*(3.40)*	*(1.50)*	*(0.78)*	*(1.64)*	*(2.47)*	*(0.81)*
J=3	0.84	0.44	0.46	0.55	0.51	0.43	0.25	0.12	0.08	0.20	0.15	0.05
	(7.71)	*(4.19)*	*(4.13)*	*(5.14)*	*(4.95)*	*(4.17)*	*(2.38)*	*(1.23)*	*(0.95)*	*(2.32)*	*(1.81)*	*(0.57)*
	Return Drift to Private Information											
J=1	-0.04	0.50	0.21	0.35	0.64	0.42	0.66	0.64	0.32	0.33	0.14	0.04
	(-0.17)	*(2.22)*	*(0.98)*	*(1.73)*	*(3.01)*	*(1.90)*	*(3.25)*	*(2.93)*	*(1.74)*	*(1.99)*	*(0.60)*	*(0.25)*
J=2	0.29	0.44	0.32	0.33	0.72	0.57	0.54	0.63	0.38	0.65	0.21	-0.34
	(1.02)	*(1.70)*	*(1.42)*	*(1.51)*	*(3.32)*	*(2.48)*	*(2.46)*	*(3.01)*	*(2.03)*	*(3.38)*	*(1.14)*	*(-1.94)*
J=3	0.33	0.39	0.59	0.61	0.56	0.57	0.70	0.56	0.71	0.40	0.04	-0.36
	(1.06)	*(1.39)*	*(2.28)*	*(2.58)*	*(2.24)*	*(2.28)*	*(3.02)*	*(2.51)*	*(3.38)*	*(2.04)*	*(0.23)*	*(-1.84)*
	Return Drift to Total Information											
J=1	0.42	1.04	0.31	0.43	0.79	0.53	0.83	0.85	0.30	0.29	0.07	0.05
	(1.60)	*(4.31)*	*(1.37)*	*(1.98)*	*(3.53)*	*(2.33)*	*(3.89)*	*(3.49)*	*(1.47)*	*(1.61)*	*(0.28)*	*(0.31)*
J=2	0.70	0.64	0.43	0.42	0.88	0.73	0.62	0.68	0.37	0.68	0.27	-0.28
	(2.36)	*(2.41)*	*(1.75)*	*(1.86)*	*(3.84)*	*(3.01)*	*(2.64)*	*(3.03)*	*(1.89)*	*(3.36)*	*(1.44)*	*(-1.56)*
J=3	0.57	0.55	0.69	0.74	0.68	0.67	0.74	0.62	0.70	0.41	0.09	-0.36
	(1.81)	*(1.93)*	*(2.51)*	*(2.98)*	*(2.59)*	*(2.55)*	*(3.01)*	*(2.63)*	*(3.25)*	*(2.06)*	*(0.48)*	*(-1.61)*

Table 2 Cont'd

Panel B: Difference in the Drift to Public Information and the Drift to Private Information

		Public Info		Private Info		Private – Public	
		profit	t-stat	profit	t-stat	profit	t-stat
J=1	M6-M1	-0.41	*(-2.31)*	0.46	*(1.52)*	0.87	*(3.26)*
	M12-M6	-0.46	*(-2.38)*	-0.38	*(-1.33)*	0.09	*(0.31)*
	M12-M1	-0.87	*(-5.05)*	0.08	*(0.26)*	0.96	*(3.40)*
	Peak-M1	-0.60	*(-2.79)*	0.71	*(2.00)*	1.31	*(4.93)*
	M12-Peak	-0.27	*(-1.37)*	-0.62	*(-2.71)*	-0.35	*(-1.31)*
J=2	M6-M1	-0.54	*(-3.96)*	0.28	*(1.05)*	0.82	*(3.52)*
	M12-M6	-0.47	*(-3.71)*	-0.91	*(-3.72)*	-0.43	*(-1.98)*
	M12-M1	-1.01	*(-7.56)*	-0.63	*(-2.13)*	0.39	*(1.64)*
	Peak-M1	-0.93	*(-6.30)*	0.34	*(1.14)*	1.27	*(4.94)*
	M12-Peak	-0.09	*(-0.73)*	-0.97	*(-3.78)*	-0.88	*(-3.69)*
J=3	M6-M1	-0.42	*(-3.03)*	0.25	*(0.83)*	0.67	*(2.48)*
	M12-M6	-0.38	*(-3.43)*	-0.94	*(-3.05)*	-0.56	*(-2.05)*
	M12-M1	-0.80	*(-5.89)*	-0.69	*(-2.19)*	0.11	*(0.43)*
	Peak-M1	-0.76	*(-5.74)*	0.38	*(1.23)*	1.14	*(3.87)*
	M12-Peak	-0.04	*(-0.31)*	-1.07	*(-3.63)*	-1.03	*(-4.18)*

Table 3: Return Drift to Short-term Information – Size Subsamples

This table reports the drift to short term total, public, and private information. i.e. formation period = 1, 2 and 3 months within each of the Large, Medium, Small subsamples according to the stock market capitalization. The results for each subsample are reported in Panel A, B and C, respectively. After forming the information portfolios, we skip one month, then hold them for twelve months. Panels tests the shape of the drift to public information and private information. Namely, we investigate the differences of the returns of the hedged portfolio in month 1, 6, 12 and the peak month when the maximum is obtained in the drift to private information. Such differences are calculated for both the drifts to public information and to private information. Then we calculate the corresponding difference-in-difference to test the difference of the shapes of the two drifts. The differences in panel B are serially correlated and we use Newey-West t-statistics to adjust for the serial correlation.

Panel A: Difference in the Drifts to Past 1-month Public and Private Information

		Public Info		Private Info		Private – Public	
		profit	t-stat	profit	t-stat	profit	t-stat
Small	M6-M1	-0.03	*(-0.08)*	0.15	*(0.32)*	0.18	*(0.30)*
	M12-M6	-0.69	*(-1.68)*	0.36	*(0.76)*	1.05	*(1.53)*
	M12-M1	-0.72	*(-1.89)*	0.51	*(1.01)*	1.23	*(2.04)*
	Peak-M1	-0.70	*(-1.73)*	0.67	*(1.31)*	1.38	*(2.25)*
	M12-Peak	-0.02	*(-0.05)*	-0.17	*(-0.47)*	-0.15	*(-0.35)*
Med	M6-M1	-0.63	*(-2.42)*	0.60	*(1.38)*	1.23	*(2.69)*
	M12-M6	-0.43	*(-1.58)*	-0.65	*(-1.60)*	-0.22	*(-0.51)*
	M12-M1	-1.05	*(-4.24)*	-0.04	*(-0.12)*	1.01	*(2.79)*
	Peak-M1	-0.38	*(-1.34)*	0.81	*(1.97)*	1.18	*(3.23)*
	M12-Peak	-0.68	*(-2.49)*	-0.85	*(-3.11)*	-0.17	*(-0.48)*
Large	M6-M1	-0.49	*(-1.78)*	0.58	*(1.43)*	1.07	*(2.40)*
	M12-M6	-0.29	*(-1.02)*	-0.72	*(-1.77)*	-0.43	*(-0.98)*
	M12-M1	-0.78	*(-2.54)*	-0.14	*(-0.32)*	0.64	*(1.41)*
	Peak-M1	-0.62	*(-1.90)*	0.73	*(1.71)*	1.35	*(2.79)*
	M12-Peak	-0.16	*(-0.58)*	-0.87	*(-2.86)*	-0.71	*(-1.89)*

Panel B: Difference in the Drifts to Past 2-month Public and Private Information

		Public Info		Private Info		Private – Public	
		profit	t-stat	profit	t-stat	profit	t-stat
Small	M6-M1	-0.86	*(-3.62)*	0.02	*(0.07)*	0.89	*(2.26)*
	M12-M6	-0.49	*(-2.43)*	-1.03	*(-4.04)*	-0.55	*(-1.71)*
	M12-M1	-1.35	*(-6.23)*	-1.01	*(-3.23)*	0.34	*(1.14)*
	Peak-M1	-1.06	*(-3.98)*	0.20	*(0.59)*	1.27	*(3.50)*
	M12-Peak	-0.29	*(-1.29)*	-1.22	*(-4.58)*	-0.93	*(-2.78)*
Med	M6-M1	-0.47	*(-2.78)*	0.43	*(1.26)*	0.90	*(3.08)*
	M12-M6	-0.65	*(-3.85)*	-1.05	*(-3.72)*	-0.40	*(-1.41)*
	M12-M1	-1.13	*(-6.05)*	-0.62	*(-1.81)*	0.50	*(1.63)*
	Peak-M1	-0.94	*(-5.24)*	0.48	*(1.33)*	1.42	*(4.37)*
	M12-Peak	-0.19	*(-1.21)*	-1.10	*(-3.85)*	-0.92	*(-3.17)*
Large	M6-M1	-0.30	*(-1.55)*	0.34	*(0.87)*	0.64	*(1.93)*
	M12-M6	-0.22	*(-1.09)*	-0.58	*(-1.69)*	-0.36	*(-1.13)*
	M12-M1	-0.52	*(-2.83)*	-0.24	*(-0.68)*	0.28	*(0.82)*
	Peak-M1	-0.59	*(-2.77)*	0.72	*(1.84)*	1.31	*(3.83)*
	M12-Peak	0.07	*(0.35)*	-0.96	*(-2.77)*	-1.03	*(-2.96)*

Panel C: Difference in the Drifts to Past 3-month Public and Private Information

		Public Info		Private Info		Private – Public	
		profit	t-stat	profit	t-stat	profit	t-stat
Small	M6-M1	-0.89	*(-5.50)*	0.14	*(0.41)*	1.04	*(3.18)*
	M12-M6	-0.32	*(-2.03)*	-1.00	*(-3.35)*	-0.68	*(-2.26)*
	M12-M1	-1.22	*(-6.92)*	-0.86	*(-2.69)*	0.36	*(1.32)*
	Peak-M1	-1.00	*(-6.28)*	0.24	*(0.63)*	1.24	*(3.35)*
	M12-Peak	-0.22	*(-1.37)*	-1.10	*(-3.66)*	-0.88	*(-3.03)*
Med	M6-M1	-0.33	*(-2.08)*	0.22	*(0.61)*	0.55	*(1.77)*
	M12-M6	-0.43	*(-3.41)*	-0.81	*(-2.42)*	-0.38	*(-1.17)*
	M12-M1	-0.76	*(-4.69)*	-0.59	*(-1.77)*	0.18	*(0.64)*
	Peak-M1	-0.60	*(-4.19)*	0.40	*(1.04)*	1.00	*(2.99)*
	M12-Peak	-0.16	*(-1.26)*	-0.99	*(-3.07)*	-0.83	*(-2.80)*
Large	M6-M1	0.22	*(1.25)*	0.31	*(0.83)*	0.08	*(0.23)*
	M12-M6	-0.33	*(-1.75)*	-0.76	*(-2.09)*	-0.43	*(-1.25)*
	M12-M1	-0.10	*(-0.71)*	-0.45	*(-1.21)*	-0.35	*(-0.97)*
	Peak-M1	0.04	*(0.19)*	0.41	*(1.05)*	0.37	*(1.06)*
	M12-Peak	-0.11	*(-0.65)*	0.47	*(1.11)*	0.59	*(1.52)*

Table 4: Return Drift to Mid- to Long-term Information

This table reports the drift to med to long-term total, public, and private information. I.e. formation period = 6, 9 and 12 months. After forming the information portfolios, we skip one month, then hold them for twelve months. In panel A, the reported numbers are the average monthly return of the hedged portfolio where we long the top information quintile and short the bottom quintile, and the corresponding t-statistics in italics. Panel B tests the shape of the drift to public information and private information. Namely, we investigate the differences of the returns of the hedged portfolio in month 1, 6 and 12. Such differences are calculated for both the drifts to public information and to private information. Then we calculate the corresponding difference-in-difference to test the difference of the shapes of the two drifts. The differences in panel B are serially correlated and we use Newey-West t-statistics to adjust for the serial correlation.

Panel A: Return Drift to Total, Public and Private Information

Month	1	2	3	4	5	6	7	8	9	10	11	12
	Return Drift to Public Information											
J=6	0.84	0.69	0.58	0.56	0.35	0.31	0.27	0.14	0.11	0.09	0.02	-0.03
	(6.80)	*(5.65)*	*(4.51)*	*(4.37)*	*(2.85)*	*(2.85)*	*(2.51)*	*(1.43)*	*(1.14)*	*(0.97)*	*(0.16)*	*(-0.34)*
J=9	0.82	0.55	0.46	0.48	0.35	0.15	0.14	0.03	-0.01	-0.01	-0.04	-0.10
	(5.63)	*(4.02)*	*(3.43)*	*(3.77)*	*(2.84)*	*(1.21)*	*(1.18)*	*(0.25)*	*(-0.06)*	*(-0.13)*	*(-0.38)*	*(-0.94)*
J=12	0.67	0.50	0.40	0.34	0.25	0.11	0.08	-0.04	-0.05	-0.01	0.02	0.12
	(4.84)	*(3.94)*	*(3.11)*	*(2.62)*	*(2.03)*	*(0.88)*	*(0.66)*	*(-0.37)*	*(-0.51)*	*(-0.12)*	*(0.17)*	*(1.21)*
	Return Drift to Private Information											
J=6	0.68	0.67	0.86	0.85	0.90	0.86	0.58	0.45	0.14	-0.02	-0.23	-0.52
	(2.12)	*(2.19)*	*(2.96)*	*(3.16)*	*(3.58)*	*(3.56)*	*(2.57)*	*(2.10)*	*(0.66)*	*(-0.08)*	*(-1.14)*	*(-2.56)*
J=9	0.93	0.89	0.99	0.81	0.75	0.48	0.22	0.13	-0.05	-0.26	-0.40	-0.57
	(2.92)	*(2.95)*	*(3.50)*	*(3.15)*	*(3.05)*	*(1.99)*	*(0.94)*	*(0.59)*	*(-0.22)*	*(-1.24)*	*(-1.91)*	*(-2.68)*
J=12	0.80	0.77	0.59	0.48	0.41	0.18	-0.05	-0.11	-0.25	-0.38	-0.49	-0.44
	(2.62)	*(2.70)*	*(2.14)*	*(1.83)*	*(1.61)*	*(0.75)*	*(-0.21)*	*(-0.49)*	*(-1.10)*	*(-1.80)*	*(-2.29)*	*(-2.10)*
	Return Drift to Total Information											
J=6	1.00	0.89	1.03	1.02	0.95	0.89	0.62	0.45	0.13	-0.03	-0.23	-0.49
	(2.94)	*(2.76)*	*(3.27)*	*(3.46)*	*(3.46)*	*(3.40)*	*(2.56)*	*(2.04)*	*(0.60)*	*(-0.14)*	*(-1.06)*	*(-2.01)*
J=9	1.22	1.13	1.07	0.93	0.78	0.48	0.24	0.16	-0.09	-0.22	-0.41	-0.57
	(3.46)	*(3.44)*	*(3.41)*	*(3.26)*	*(2.95)*	*(1.80)*	*(0.95)*	*(0.65)*	*(-0.39)*	*(-0.99)*	*(-1.89)*	*(-2.27)*
J=12	1.04	0.94	0.63	0.54	0.44	0.16	-0.08	-0.13	-0.31	-0.39	-0.43	-0.39
	(3.09)	*(3.04)*	*(2.05)*	*(1.90)*	*(1.60)*	*(0.62)*	*(-0.32)*	*(-0.52)*	*(-1.31)*	*(-1.68)*	*(-1.88)*	*(-1.74)*

Panel B: Difference in the Drift to Public Information and the Drift to Private Information

		Public Info		Private Info		Private – Public	
		profit	t-stat	profit	t-stat	profit	t-stat
J=1	M6-M1	-0.54	*(-3.77)*	0.18	*(0.71)*	0.71	*(2.77)*
	M12-M6	-0.34	*(-2.30)*	-1.38	*(-4.08)*	-1.04	*(-3.92)*
	M12-M1	-0.88	*(-6.30)*	-1.20	*(-3.75)*	-0.32	*(-1.26)*
J=2	M6-M1	-0.67	*(-4.08)*	-0.45	*(-1.23)*	0.22	*(0.78)*
	M12-M6	-0.24	*(-1.50)*	-1.05	*(-3.03)*	-0.81	*(-3.01)*
	M12-M1	-0.91	*(-5.98)*	-1.50	*(-4.70)*	-0.59	*(-2.35)*
J=3	M6-M1	-0.56	*(-3.49)*	-0.62	*(-1.70)*	-0.06	*(-0.21)*
	M12-M6	0.02	*(0.10)*	-0.62	*(-1.80)*	-0.64	*(-2.55)*
	M12-M1	-0.55	*(-3.21)*	-1.24	*(-3.42)*	-0.69	*(-2.52)*