Strategic informed trading and the market reaction to earnings announcements[#]

Kee H. Chung*
State University of New York (SUNY) at Buffalo and
Sungkyunkwan University

Oliver Kim
University of Maryland

Steve C. Lim
Texas Christian University

Sean Yang
State University of New York (SUNY) at Buffalo

February 10, 2016

[#] The authors thank Stephen Brown, Rick Cazier, David Hirshleifer, Sahn-Wook Huh, Jack Jiang, Robert Kim, Sangwan Kim, Michael Kimbrough, Larry Lockwood, David Ng, Claudia Zhen Qi, Steven Sherwood, Inho Suk, Hongping Tan, Cristian Tiu, Cihan Uzmanoglu, Bob Vigeland, Zhichen Wang, William Wempe, Mao Ye, Woongsun Yoo, Lingyin Zhu, session participants at the 2014 Northern Finance Association conference, the 2014 Financial Management Association conference, the 24th Annual Conference on Financial Economics and Accounting, the 8th Annual Lone Star Accounting Research Conference, the 2013 AAA Conference, Mid-Atlantic AAA Conference, Western AAA Conference, and seminar participants at Chinese University of Hong Kong, City University of Hong Kong, KAIST Graduate School of Finance & Accounting, Texas Christian University, University of Maryland, and University of Texas at Arlington for useful comments and discussions. An earlier version of this paper was circulated under the title "Information asymmetry, information content, and the measurement of underreaction to earnings announcements." This work is supported by the National Research Foundation of Korea Grant funded by the Korean Government (NRF-2014S1A3A2036037). All errors are solely the authors' responsibility.

*Corresponding Author: Kee H. Chung, School of Management, State University of New York at Buffalo, Buffalo, NY 14260; Tel.: (716) 645-3262; E-mail address: keechung@buffalo.edu.

Strategic informed trading and the market reaction to earnings announcements

Abstract

We show that strategic informed trading that arises from the information asymmetry between the liquidity demander and the liquidity provider results in underreaction to earnings announcements. Specifically, we show that $ERC = k \cdot (1-\rho^2)$ and $PEAD = k \cdot \rho^2$, where ERC is the earnings response coefficient, $PEAD$ is the post-earnings announcement drift, k is the information content of earnings, and ρ is the correlation coefficient between order imbalance and earnings surprise. In our model, ρ^2 is an empirical metric of the information asymmetry-driven underreaction that can be used to predict the size of the post-earnings announcement drift. We provide strong empirical evidence that is consistent with our analytical predictions.

JEL Classification Code: G14

Key words: Strategic trading, Information asymmetry, Information precision, Liquidity demander, Liquidity provider, Order imbalance, Information content, Price impact

1. Introduction

In this study we analyze how strategic informed trading that arises from the information asymmetry between the liquidity demander and the liquidity provider affects both the immediate market reaction to earnings announcements and the post-earnings announcement drift (*PEAD*).[1] Although there is an extensive literature analyzing the immediate and delayed market reactions to earnings announcements, prior research does not *explicitly* recognize in an analytical framework that the market reactions to earnings announcements depend on strategic informed trading.[2] Our study contributes to the literature by developing an analytical model that relates strategic informed trading to the post-earnings announcement drift and testing its empirical predictions.

Prior research suggests that earnings announcements increase information asymmetries among traders. Kim and Verrecchia (1994) maintain that earnings announcements provide information that allows some traders to make better judgments about a firm's performance than others, increasing information asymmetries among traders during the earnings announcement period. Subsequent studies find evidence that is generally consistent with this prediction.[3] Kaniel, Liu, Saar, and Titman (2012) show that the private-information-based trading by individual investors around earnings announcements accounts for about half of the abnormal returns on and after earnings announcement dates. We incorporate these results into an analytical model and show that the degree of market underreaction to earnings announcements increases with the information asymmetry between the liquidity demander and the liquidity provider.

Prior studies propose a number of possible explanations for why and how *PEAD* occurs.

---

[1] In our analysis, this information asymmetry results from the *underutilization* of earnings information by the liquidity provider.

[2] See, e.g., Kothari (2001) and Zhang, Cai, and Keasey (2013).

[3] See, e.g., Lee, Mucklow, and Ready (1993), Coller and Yohn (1997), Yohn (1998), and Bhat and Jayaraman (2009).

One strand of research provides behavioral explanations of *PEAD*. For instance, the investor inattention hypothesis predicts that the market response to earnings announcement is muted and *PEAD* is higher on days of investor inattention (Hirshleifer et al., 2009; Hirshleifer et al., 2011; Mian and Sankaraguruswamy, 2012). The disposition effect theory says that investors hold on to losing stocks for too long and sell winning stocks too soon, generating stock price underreactions to news and subsequent *PEAD* (Shefrin and Statman, 1985; Bernard and Thomas, 1989 and 1990; Odean, 1998; Frazzini, 2006; Barberis and Xiong, 2009).

Another strand of research explains *PEAD* based on trading costs and information uncertainty. For example, the transaction cost hypothesis posits that trading costs deter informed traders from aggressively arbitraging pricing errors at the time of earnings announcements (Bhushan, 1994; Ke and Ramlingegowda, 2005; Sadka, 2006; Ng et al., 2008; Chordia et al., 2009). The information risk hypothesis explains that information risk or information uncertainty deters investors from reacting fully to the information in earnings announcements, creating initial underreaction and subsequent *PEAD* (Mendenhall, 2004; Zhang, 2006; Garfinkel and Sokobin, 2006; Francis et al., 2007, Zhang, Cai, and Keasey, 2013). In the present study, we extend the latter strand of research by providing an analytical model of *PEAD* and testing the unique empirical implications of the model that have not been explored in prior research.[4]

A common explanation for persistent mispricing in the securities market is the limits of arbitrage. That is, the arbitrage trading corrects the mispricing only up to the point where the marginal benefit justifies the cost of arbitrage trading (e.g., Grossman and Stiglitz, 1980; Shleifer and Vishny, 1997). Therefore it is the cost of arbitrage, *not* the intensity of arbitrage

[4] Although prior studies have examined the effect of information asymmetry and information risk on the market reaction to earnings announcements, there has been no formal analytical treatment of the issue. As a result, prior research (e.g., Zhang, Cai, and Keasey, 2013) does not make a clear distinction between fundamental (information) uncertainty and information asymmetry. Our study sheds further light on the issue using an analytical model.

trading, that determines the magnitude of the residual mispricing (e.g., *PEAD*). This line of thought assumes that the market is perfectly competitive in the long run and the marginal arbitrageur is indifferent between trading and non-trading.

However, the market does not always become perfectly competitive soon after the firm makes a public announcement. This is because the "complete" extraction of the value-relevant information in a public announcement requires efforts, skills, and time, and is costly for most of market participants (Hirshleifer and Teoh, 2003). If only a few investors were able to immediately and completely extract the pricing implication of the announced information, they would capitalize on it by strategically exercising their trading options (Kyle, 1985). That is, they would optimize their trading by taking into account of its impact on the correction of mispricing. This "strategic trading" is exactly what generates the *slow* price adjustment, or a Kyle-type post-announcement drift and this mechanism gives rise to an association between the intensity of strategic trading and the stock price underreaction.

To incorporate information asymmetry in an analytical framework, we employ a model in which some market participants have better information processing skills than others, which gives rise to information asymmetry between market participants. The market participants who possess better information processing skills have an information advantage and strategically exploit the advantage through trading. We invoke the assumption that some market participants better utilize earnings information than others from the fact that (1) a typical earnings announcement contains a large amount of information and data that are subject to different interpretations and (2) there are multiple sources of earnings forecasts with different information.[5] The sheer size and diverse sources of information about earnings are likely to result in information asymmetries between market participants.

[5] See, for example, Apple's financial results for its fiscal 2014 first quarter ending on December 28, 2013 available at the following website: http://www.apple.com/pr/library/2014/01/27Apple-Reports-First-Quarter-Results.html. There were more than 30 analysts providing earnings forecasts for Apple in the first quarter of 2014.

Our model has three types of traders: the liquidity provider, the liquidity demander, and the utilitarian trader. The liquidity provider furnishes liquidity (immediacy) to the other traders by offering a price quote based on partial earnings information and the total order received from both the liquidity demander and the utilitarian trader. The liquidity demander strategically submits (market) orders based on full earnings information and trades at the quote set by the liquidity provider. Finally, the utilitarian trader trades for non-informational reasons.

We define the information asymmetry between the liquidity demander and the liquidity provider as the ratio between the precision of the information used by only the liquidity demander and the precision of the information used by both the liquidity demander and the liquidity provider. Hence, information asymmetry increases with the precision of the information exclusively used by the liquidity demander and decreases with the precision of the information used by both the liquidity demander and provider.

In our model, price impounds earnings information in two different ways. Price fully and instantly impounds the earnings information used by both the liquidity demander and the liquidity provider. This portion of earnings information is unrelated to either the liquidity demander's order size or the total order imbalance because no market participants can generate a profit from this information. The portion of earnings information used only by the liquidity demander, however, gets into price through his strategic demand order. If the news is good (bad) the liquidity demander submits a buy (sell) order, and the liquidity provider raises (lowers) his price quote upon receiving the order.[6] This portion of earnings information gets into price only *partially* because the liquidity provider cannot *perfectly* infer the earnings information used exclusively by the liquidity demander from the total order received. When there is a higher

[6] Lee (1992) shows that quarterly earnings announcements with good (bad) news relative to the *Value Line Investment Survey* forecast trigger brief, but intense, buying (selling) in large trades. Bhattacharya (2001) and Battalio and Mendenhall (2005) show that small traders tend to trade in response to random-walk earnings forecast errors while large traders tend to trade in response to analysts' earnings forecast errors.

degree of information asymmetry, a larger portion of earnings information gets into price through the liquidity demander's order, and thus a smaller portion of earnings information gets into price at the time of announcement (i.e., a smaller earnings response coefficient (*ERC*)) and a larger portion gets into price during the post-earnings announcement period (i.e., a larger *PEAD*).

We define the degree of market underreaction to earnings announcements as the proportion of the earnings information that the stock price fails to impound instantly. In our model, underreaction increases with the information asymmetry between the liquidity demander and the liquidity provider but decreases with the information content of earnings. The information content of earnings in our model is the *total* stock price reaction to earnings announcement, and it increases with the precision of the earnings information released at the announcement relative to the precision of prior belief about earnings.

The liquidity demander's strategic exploitation of information advantage leads to a positive correlation coefficient (ρ) between order imbalance and earnings surprise. As this information asymmetry increases, more earnings information gets into price through the liquidity demander's order and thus ρ increases. We show that ρ^2 is the degree of underreaction and use this metric as our theory-based empirical measure of underreaction. Since underreaction is the cause of *PEAD*, ρ^2 is a predictor of *PEAD*, and the size of *PEAD* depends on both ρ^2 and the information content of earnings. Specifically, we show that $PEAD = k \cdot \rho^2$, where *k* is the information content of earnings.

Our main empirical findings could be summarized as follows. We find a large positive (negative) order imbalance during the two-day event window when earnings news is good (bad), suggesting the liquidity demander's exploitation of information advantage. To further confirm this result, we test the relation between a model-implied information asymmetry measure and various firm/stock attributes that prior research has used to measure information asymmetry around corporate events. We show that our model-implied information asymmetry measure is significantly related to these attributes in the predicted manner, a result from which we gain confidence in the validity of our model.

Consistent with the predictions of our analytical model, we show that the observed degree of underreaction is positively related to the model-implied information asymmetry measure, negatively related to the estimate of the information content of earnings, and positively related to our empirical metric of underreaction (ρ^2). We show that both our model-implied information asymmetry measure and ρ^2 explain *PEAD* and their explanatory power does not decrease even after we control for other information asymmetry proxies used in prior research.

Our empirical results show that the explanatory power of our model-implied information asymmetry measure and ρ^2 is higher than that of firm attributes (i.e., the bid-ask spread, opinion divergence, firm size, analyst following, and institutional ownership) that prior studies used to explain the post-earnings announcement drift. We interpret this result as evidence that our theory-based constructs of information asymmetry and underreaction provide a more accurate measurement of underreaction (thus, a more accurate prediction of *PEAD*) than these *ad hoc* measures of the firm's information environment.

Consistent with our analytical prediction, we also find that the relation between *PEAD* and ρ^2 is steeper when the information content of earnings is larger. Finally, we show that a trading strategy based on both ρ^2 and the estimate of information content is also profitable. Using the Pastor-Stambugh (2003) four-factor alpha, we show that the annual profit is as large as 11.64%. Out of this profit, around 8.25% is attributable to ρ^2 and 3.39% is attributable to the information content of earnings.

To the best of our knowledge, the present study is the first that measures underreaction to earnings announcements using a theory-based empirical metric of underreaction. Another important contribution of our study is that we provide a unifying framework for analyzing the market reaction to earnings announcements. A number of previous studies analyze the determinants of *ERC*.[7] There is also a large literature that tries to understand the causes and consequences of the post-earning announcement drift (*PEAD*). However, there has been little attempt to understand the interrelatedness of *ERC* and *PEAD* within a unified analytical framework or empirical analysis. The joint analysis of *ERC* and *PEAD* should prove useful because *ERC* and *PEAD* span the *total* information content of earnings. Our study provides such a framework, along with pertinent empirical evidence.

We organize the rest of the paper as follows. Section 2 presents our model and derives analytical expressions for *ERC*, *PEAD*, and ρ. Section 3 develops our testable hypotheses from the model. Section 4 explains our data sources, variable measurements, and descriptive statistics. Section 5 tests our hypotheses. Section 6 evaluates the explanatory power of ρ^2 and our empirical metric of information asymmetry relative to that of other firm/stock attributes used in prior research to explain the post-earnings announcement drift. Section 7 assesses the profitability of trading strategies that are based on the empirical metrics developed in our paper. Finally, we provide a brief summary and concluding remarks in Section 8.

2. The model and analytical results

2.1. The model

In this section, we introduce our analytical model in which we define information asymmetry and the information content of earnings. There are three types of traders in our model: the liquidity demander who fully utilizes earnings information at the time of earnings announcements, the liquidity provider who posts price quotes based on the partial earnings

[7] See, e.g., Collins and Kothari (1989).

information, and the utilitarian trader who trades for non-informational reasons.

In the real market setting, both limit-order traders and market makers play the role of the liquidity provider in our model.[8] They establish their price quotes based on the information available to them, and other traders buy and sell at their quotes. Traders who consume (or demand) liquidity by submitting market orders play the role of the liquidity demander in the model. They are aggressive traders who typically have informational advantages over other market participants (e.g., limit order traders and market makers) and strategically trade at prices set by the liquidity provider. Lastly, those traders who submit buy or sell orders for non-informational reasons play the role of the utilitarian trader in our model.

Consider a market with one risky asset (the firm) and a riskless bond. The final payoff from one unit of the risky asset is $\tilde{u}$, which is normally distributed with mean 0 and precision (inverse of variance) *h*. There are three time periods. Period 1 is the pre-earnings-announcement period in which no traders possess any information other than the prior distribution of $\tilde{u}$. Period 2 is the earnings announcement period. The critical assumption in this model is that the earnings signal $\tilde{x}$, which arrives in period 2, consists of two noisy information signals, $\tilde{y}$ and $\tilde{z}$. Signal $\tilde{y} = \tilde{u} + \tilde{\eta}$ is a signal used by both the liquidity provider and the liquidity demander and $\tilde{z} = \tilde{u} + \tilde{\varepsilon}$ is a signal used only by the liquidity demander, where $\tilde{\eta}$ and $\tilde{\varepsilon}$ are both normally and independently distributed with mean zero and precision *m* and *s*, respectively. One may interpret $\tilde{y}$ and $\tilde{z}$ as two imperfect understandings (or knowledge) in period 2 on the pricing implication of the announced earnings, and $\tilde{x}$ as the one-signal summary of $\tilde{y}$ and $\tilde{z}$ written as precision-weighted knowledge:

[8] Limit order traders play a critical role in the price discovery process on the NYSE and NASDAQ during our study period. The U.S. Securities and Exchange Commission introduced the limit order display rule in 1997 that requires market makers and specialists to reflect public limit orders in their quotes when the orders are better than their own quotes. This rule ensures that the general public competes directly with market makers and specialists in the price discovery process.

$$\tilde{x} = \frac{m}{m+s}\tilde{y} + \frac{s}{m+s}\tilde{z}.$$

A typical earnings report comprises a large amount of information and multiple forecasts usually precede each earnings announcement. In the context of our model, $\tilde{x}$ could be viewed as a composite measure of *earnings surprise* (i.e., actual – predicted) implied by all the information contained in both announced earnings and available forecasts of earnings. Our model assumes that the liquidity demander extracts information $\tilde{y}$ and $\tilde{z}$ from all the sources and utilizes them in trading, while the liquidity provider extracts only information $\tilde{y}$.

In period 3 the final payoff $\tilde{u}$ is determined and consumption occurs. In period 2, the liquidity demander submits an order of size $\tilde{d}$ to the liquidity provider based on his available information. The utilitarian trader also submits an order of size $\tilde{\ell}$ to the liquidity provider based on his random needs. We assume that $\tilde{\ell}$ is normally and independently distributed with mean 0 and variance *L*. The two orders are batched together and the liquidity provider only observes the *order imbalance* $\tilde{\omega} = \tilde{d} + \tilde{\ell}$. The liquidity provider conjectures that the liquidity demander's order $\tilde{d}$ is a linear function of the two information signals in the form of $\tilde{d} = \beta\tilde{y} + \gamma\tilde{z}$ and sets the price as the expectation of $\tilde{u}$ conditional on the available information at the time. That is, $\tilde{P}_2 = E[\tilde{u} \mid \tilde{y}, \tilde{\omega}]$. The liquidity demander strategically decides his order based on a conjecture that the price chosen by the liquidity provider is a linear function of the information available to the liquidity provider at the time. That is, we can express the liquidity demander's conjecture as $\tilde{P}_2 = \alpha\tilde{y} + \lambda\tilde{\omega}$.

In this model it is straightforward to show that there is a unique equilibrium in which:

$$\tilde{P}_1 = 0,$$

$$\alpha = \frac{m}{h+m},$$

$$\lambda = \frac{1}{2}\sqrt{\frac{s}{(h+m)(h+m+s)L}},$$

$$\beta = -\frac{ms}{2\lambda(h+m)(h+m+s)} = -\sqrt{\frac{m^2 sL}{(h+m)(h+m+s)}},$$

and

$$\gamma = \frac{s}{2\lambda(h+m+s)} = \sqrt{\frac{(h+m)sL}{h+m+s}}.$$

Thus, the liquidity demander's order size and the liquidity provider's price are determined as

$$\tilde{d} = \frac{s}{2\lambda(h+m+s)} \cdot \left[\tilde{z} - \frac{m}{h+m}\tilde{y}\right]$$

and

$$\tilde{P}_2 = \frac{m}{h+m}\tilde{y} + \lambda(\tilde{d} + \tilde{\ell})$$

$$= \frac{m(2h+2m+s)}{2(h+m)(h+m+s)}\tilde{y} + \frac{s}{2(h+m+s)}\tilde{z} + \frac{1}{2}\sqrt{\frac{s}{(h+m)(h+m+s)L}}\tilde{\ell}, \qquad (1)$$

respectively. The liquidity demander's order $\tilde{d}$ is proportional to the difference between $\tilde{z}$, which only the liquidity demander uses, and the liquidity provider's expectation of $\tilde{z}$, $\frac{m\tilde{y}}{h+m}$, based on his information. As a result, we have

$$Cov(\tilde{y}, \tilde{\omega}) = Cov(\tilde{y}, \tilde{d} + \tilde{\ell}) = Cov(\tilde{y}, \tilde{d})$$

$$\propto Cov(\tilde{y}, \tilde{z} - \frac{m}{h+m}\tilde{y}) = Cov(\tilde{u} + \tilde{\eta}, \frac{h\tilde{u} - m\tilde{\eta}}{h+m}) = 0. \qquad (2)$$

That is, in equilibrium the liquidity demander's order size and also the total order imbalance are not based on, and thus are not correlated with, the earnings signal used by both the liquidity demander and provider.

We define information asymmetry between the liquidity demander and the liquidity provider as the ratio between the precision (s) of the information ($\tilde{z}$) used by only the liquidity demander and the precision (s + m) of the liquidity demander's total information ($\tilde{z}$ and $\tilde{y}$), i.e., $\theta \equiv \frac{s}{m+s}$. Information asymmetry increases with the precision (s) of the information exclusively used by the liquidity demander and decreases with the precision (m) of the information used by both the liquidity demander and provider.

Finally, consider regressing the final payoff, $\tilde{u}$, on earnings surprise, $\tilde{x}$. We define the information content of earnings (i.e., informativeness of $\tilde{x}$) as the regression coefficient on $\tilde{x}$, i.e., $k \equiv \frac{Cov[\tilde{u},\tilde{x}]}{Var[\tilde{x}]}$. When the sum of precisions (m + s) of the information released at the time of earnings announcement is high, $\tilde{x}$ is close to $\tilde{u}$ and k is close to 1. On the other hand, when the precision is low, $\tilde{x}$ is not informative and k is close to 0. In Appendix A we show that $k \equiv \frac{Cov[\tilde{u},\tilde{x}]}{Var[\tilde{x}]} = \frac{m+s}{h+m+s}$. Note that h is an inverse measure of the inherent (or fundamental) uncertainty in earnings, which is likely to be determined by the firm's business and operating risks. If the firm's business and operating risks were low (i.e., h is large), the information content of earnings would be low even if its precision (i.e., m + s) is high because there is little uncertainty in the final payoff to begin with. In contrast, if the firm's business and operating risks were high (i.e., h is small), the information content of earnings would be high even if its precision is low.

2.2. ERC and PEAD as functions of information content and information asymmetry

In our model we introduce strategic informed trading as a new property of market reaction to an earnings announcement independent from the information content of earnings. Thus, the two factors, information asymmetry and information content, determine the impact of the earnings announcement on share price and *PEAD*. Consider regressing the price change in the earnings announcement period, $\tilde{P}_2 - \tilde{P}_1 = \tilde{P}_2$, on earnings surprise, $\tilde{x}$. We derive in

Appendix A the following result for the earnings response coefficient (*ERC*):

$$ERC[\tilde{x}] = \frac{Cov[\tilde{P}_2, \tilde{x}]}{Var[\tilde{x}]} = \frac{2m(h+m+s)+hs}{2(h+m)(h+m+s)}$$

$$= k \cdot \frac{2-(1+k)\theta}{2-2k\theta}. \tag{3}$$

In the last expression above, we use two ratio parameters, information content (k) and information asymmetry (θ).[9] If there is no information asymmetry between the liquidity provider and the liquidity demander, i.e., if $\theta = 0$, *ERC* is equal to the information content of earnings, k. If there is information asymmetry between the liquidity provider and the liquidity demander, i.e., if $\theta > 0$, *ERC* is smaller than k. In the extreme case of $\theta = 1$, i.e., if the entire earnings information is asymmetrically utilized, *ERC* is equal to $\frac{k}{2}$.

From equation (3) we have two comparative static results: $\frac{\partial ERC}{\partial k} > 0$ and $\frac{\partial ERC}{\partial \theta} < 0$. That is, *ERC* increases with the information content of earnings and decreases with information asymmetry. The following proposition summarizes these results:

Proposition 1: $ERC \equiv \frac{Cov[\tilde{P}_2, \tilde{x}]}{Var[\tilde{x}]} = k \cdot \frac{2-(1+k)\theta}{2-2k\theta}$ *is increasing in k and decreasing in* θ*, where k is the information content of earnings, and* θ *is the information asymmetry between the liquidity demander and the liquidity provider.*

From the definition of k (i.e., $k = \frac{m+s}{h+m+s}$), the information content (k) of earnings announcements increases with information risk (inverse of h). The positive relation between *ERC* and k is consistent with the information content hypothesis of Zhang, Cai, and Keasey (2013) that information risk reduces the informativeness of price, raises the relative importance

[9] The range of *ERC* is between 0 and 1 in our one period model. In a model where an earnings number represents a permanent stream, the *ERC* would be multiplied by $1+1/r$, where r is the discount rate.

of earnings announcements in the price discovery process, and thus increases the initial market reaction per unit of earnings surprise. Our study provides further insight that the information asymmetry between the liquidity demander and the liquidity provider is another determinant of the initial market reaction.

Now consider regressing the price change during the post-earnings announcement period, $\tilde{u} - \tilde{P}_2$, on earnings surprise, $\tilde{x}$. We derive (in Appendix B) the following result for the post-earnings announcement drift (*PEAD*):

$$PEAD \equiv \frac{Cov[\tilde{u} - \tilde{P}_2, \tilde{x}]}{Var[\tilde{x}]} = k - ERC = k \cdot \frac{(1-k)\theta}{2(1-k\theta)}. \tag{4}$$

From equation (4), when the liquidity demander does not have an information advantage (i.e., $\theta = 0$), *PEAD* is zero. In this case, *ERC* reflects the entire information content of earnings (i.e., *ERC* = *k*) and thus there is no residual price adjustment after the earnings announcement period (i.e., *PEAD* = 0). When all earnings information gets into price through the liquidity demander's order (i.e., $\theta = 1$ and *ERC* $= \frac{k}{2}$), *PEAD* is equal to $\frac{k}{2}$. It is easy to show that $\frac{\partial PEAD}{\partial \theta} > 0$. The partial derivative of *PEAD* with respect to *k*, however, is not monotonic. We find that $\frac{\partial PEAD}{\partial k} > 0$ if $k(2 - k\theta) < 1$, but $\frac{\partial PEAD}{\partial k} < 0$ if $k(2 - k\theta) > 1$. The following proposition summarizes these results:

Proposition 2: $PEAD \equiv \frac{Cov[\tilde{u} - \tilde{P}_2, \tilde{x}]}{Var[\tilde{x}]} = k \cdot \left[\frac{\theta}{2} \frac{1-k}{(1-k\theta)}\right]$ *is increasing in θ, where k is the information content of earnings and θ is information asymmetry between the liquidity demander and the liquidity provider. PEAD is increasing in k if* $k(2 - k\theta) < 1$, *but decreasing in k if* $k(2 - k\theta) > 1$.

It is useful to characterize the degree of market underreaction (*DMU*) to earnings announcement as the proportion of the information content of earnings that stock price fails to impound instantly. From equation (4), it is straightforward to show that this proportion is *DMU* $= \frac{\theta}{2}\frac{1-k}{(1-k\theta)}$. Note from equation (3) that $ERC = k\cdot\left[1-\left(\frac{\theta}{2}\frac{1-k}{(1-k\theta)}\right)\right] = k\cdot(1-DMU)$. Hence, $1-\left(\frac{\theta}{2}\frac{1-k}{(1-k\theta)}\right)=1-DMU$ represents the proportion of the information content of earnings that is incorporated into stock price at the time of earnings announcement.

A simple comparative statics analysis shows that the degree of underreaction (*DMU*) is increasing in θ but decreasing in *k*. We show in Figure 1a and Figure 1b the relation between the degree of underreaction and *k* at different levels of θ and the relation between the degree of underreaction and θ at different levels of *k*, respectively.[10] The positive relation between the degree of underreaction and θ shown in Figure 1b is generally steeper than the negative relation between the degree of underreaction and *k* for most values of *k* shown in Figure 1a. These results suggest that variation in the degree of underreaction is more likely driven by variation in information asymmetry than by variation in information content. The following corollary summarizes these results.

Corollary 1: *The degree of market underreaction to earnings announcement,* $\frac{\theta}{2}\frac{1-k}{(1-k\theta)}$, *is increasing in* θ *but decreasing in k, where* θ *is the information asymmetry between the liquidity demander and the liquidity provider and k is the information content of earnings.*

2.3. Correlation between order imbalance and earnings surprise (ρ *)*

Consider the correlation coefficient between order imbalance and earnings surprise

[10] The maximum possible value for the degree of underreaction is 0.5, and this is the case when θ is equal to 1.

(ρ).[11] We show in equation (2) that the correlation between order imbalance ($\tilde{\omega}$) and the part of earnings information ($\tilde{y}$) that both the liquidity demander and the liquidity provider use is zero. In contrast, we expect a positive correlation between order imbalance and the earnings information ($\tilde{z}$) used only by the liquidity demander because the intensity and direction of the information-based trading (and thus $\tilde{\omega}$) are a function of $\tilde{z}$. Because $Corr(\tilde{\omega}, \tilde{y}) = 0$, we expect the correlation between order imbalance ($\tilde{\omega}$) and the entire earnings information (i.e., $\tilde{x} = (1-\theta)\tilde{y} + \theta\tilde{z}$) to capture the extent to which a subset of market participants (i.e., the liquidity demander in our model) exploit the information advantage and the resulting information asymmetry among traders.

We show in Appendix A that the correlation coefficient between $\tilde{x}$ and $\tilde{\omega}$ can be expressed as:

$$\rho \equiv Corr[\tilde{x}, \tilde{\omega}] = \frac{Cov[\tilde{x}, \tilde{\omega}]}{\sqrt{Var[\tilde{x}] \cdot Var[\tilde{\omega}]}} = \sqrt{\frac{hs}{(h+m)(m+s)}} = \sqrt{\frac{\theta - k\theta}{2(1-k\theta)}}. \tag{5}$$

Note that equation (5) can be rewritten as follows:

$$\rho^2 = \frac{\theta(1-k)}{2(1-k\theta)}, \tag{6}$$

and from equations (3) and (4) we have $ERC = k \cdot (1-\rho^2)$ and $PEAD = k \cdot \rho^2$. That is, the degree of market underreaction is simply ρ^2.

Because ρ^2 is the coefficient of determination (R^2) in the regression of order imbalance on earnings surprise, the degree of underreaction is equal to the proportion of the variation in order imbalance that is due to the variation in earnings surprise. To see the economic intuition behind this result, note first that underreaction would occur only if there

[11] Chordia, Roll, and Subrahmanyam (2002) and Chordia and Subrahmanyam (2004) analyze the relation between order imbalance and stock returns.

were informed trading because the liquidity provider fails to fully incorporate all information into price. If there were no informed trading, there would be no underreaction, and order imbalance would also be unrelated to earnings surprise because the uninformed traders' order sizes ($\beta\tilde{y}$ and $\tilde{\ell}$) are unrelated to earnings surprise. Hence, no underreaction would be observed when there is zero correlation between order imbalance and earnings surprise. As informed trading increases, the degree of underreaction and the correlation between order imbalance and earnings surprise would increase simultaneously. These considerations suggest why the degree of underreaction increases with the proportion of the variation in order imbalance that is due to the variation in earnings surprise.

The following proposition summarizes the above results:

Proposition 3: $ERC = k \cdot (1-\rho^2)$, $PEAD = k \cdot \rho^2$, and $r^2 = \dfrac{\theta(1-k)}{2(1-k\theta)}$, *where* ρ^2 *is the squared correlation coefficient between order imbalance and earnings surprise, k is the information content of earnings, and θ is the information asymmetry between the liquidity demander and the liquidity provider*.

Although there is a large body of literature analyzing the effect of earnings announcements on share prices, prior research does not explicitly recognize that the post-earnings announcement drift depends not only on the information content of earnings (k) but also on ρ^2, which is largely determined by the information asymmetry (θ) between traders.[12] For instance, Zhang, Cai, and Keasey (2013) examine the effects of information risk and transaction costs on *ERC* and *PEAD*. They show that the earnings announcements of firms with higher levels of information risk (i.e., lower h in our model) convey more information and thus result in higher *ERC* and that higher transaction costs decrease *ERC* and increase *PEAD*. Our

[12] We showed earlier in Figure 1a and Figure 1b that although ρ is a function of both information content (k) and information asymmetry (θ), ρ is largely determined by information asymmetry.

study differs from theirs in that earnings announcements invoke information asymmetry in our study whereas earnings announcements provide new information that reduces information risk in their study. Proposition 3 shows how to disentangle the effects of these two properties (i.e., information content and information asymmetry) of earnings announcements on the post-earnings announcement drift.

3. Testable hypotheses

In this section we develop testable hypotheses based on the results in Section 2.

3.1. Measurement of information asymmetry and its empirical determinants

Our measure of information asymmetry (θ) is not directly observable because it is a function of two unobservable variables m and s, where m denotes the precision of the information used by both the liquidity demander and provider and s denotes the precision of the information exclusively used by the liquidity demander. However, Proposition 3 enables us to estimate θ by expressing θ as a function of two empirically measurable variables:

$$\hat{\theta} = \frac{2\hat{\rho}^2}{1 - \hat{k} + 2\hat{k} \cdot \hat{\rho}^2}, \tag{7}$$

where $\hat{\theta}$ is the model-implied information asymmetry measure, $\hat{\rho}$ is the empirical estimate of the correlation coefficient between order imbalance and earnings surprise, $\hat{k}$ is the coefficient on $\tilde{x}$ in the regression of the final payoff ($\tilde{u}$) on earnings surprise ($\tilde{x}$) (see Section 5 for details).

If θ measures the liquidity demander's information advantage, then we expect $\hat{\theta}$ to be associated with the firm/stock attributes that prior research has used to proxy for the information asymmetry around corporate events (e.g., Huang and Stoll, 1996; Easley, Kiefer, O'Hara, and Paperman, 1996; Garfinkel and Sokobin, 2006). These firm/stock attributes include the price impact of a trade (measured by the adverse selection component of the spread), the effective bid-ask spread, the probability of informed trading (*PIN*), the divergence of investors' opinions, firm size, the number of analysts, and the number of institutional investors.

A large price impact and wide effective bid-ask spread both indicate a high degree of information asymmetry. Hence, we expect $\hat{\theta}$ to be positively related to these variables measured at the time of earnings announcement. A high *PIN* value also indicates a high degree of information asymmetry. Different from the previous two variables, *PIN* captures the extent of informed trading before earnings announcement. Because θ measures the pre-existing information asymmetry across investors before earnings announcement and high ρ is simply a manifestation of this pre-existing information asymmetry, we hypothesize that $\hat{\theta}$ increases with the degree of informed trading (*PIN*) before earnings announcement.[13]

Garfinkel and Sokobin (2006) use the divergence of investors' opinions as a measure of the pre-existing information asymmetry before earnings announcement and show that a high pre-existing information asymmetry leads to a low share price at the time of earnings announcement and a positive price drift afterwards. Similarly, we hypothesize that $\hat{\theta}$ increases with the divergence of investors' opinions. Prior research (e.g., Brennan and Subramanyam, 1995; Atkins et al., 2012) suggests that larger firms, firms followed by more analysts, and firms with more institutional investors are associated with a lower degree of information asymmetry. Thus, we expect $\hat{\theta}$ to be negatively related to firm size, the number of analysts following the firm, and the number of institutions that hold the firm's shares. These considerations lead to the following hypothesis.

Hypothesis 1: *The model-implied information asymmetry measure (i.e.,* $\hat{\theta}$ *) is positively*

[13] Vega (2006) shows that high *PIN* is associated with insignificant post-earnings announcement drift because high informed trading before earnings announcement is associated with low rational structural uncertainty at the time of earnings announcement.

related to price impact, effective spread, PIN, and the divergence of investors' opinions and negatively related to firm size, number of analysts, and number of institutional investors.

3.2. Determinants of underreaction

From Corollary 1 and Proposition 3, the degree of underreaction (ρ^2) is positively associated with θ and negatively associated with *k*. Although we do not expect that the *observed* degree of underreaction is exactly equal to our estimate of ρ^2 because underreaction could be driven by many other factors that are not considered in our analytical model, we expect the observed degree of underreaction to be positively related to our estimate of ρ^2. These considerations lead to the following hypothesis:

Hypothesis 2: *The observed degree of underreaction to earnings announcement is positively related to information asymmetry (θ), negatively related to the information content of earnings (k), and positively related to the square of the correlation between order imbalance and earnings surprise (ρ^2).*

3.3. Predictability of PEAD based on ρ^2, k, and θ

Proposition 2 shows that *PEAD* is a positive function of θ. Proposition 3 shows that *PEAD* is the product of *k* and ρ^2. Proposition 3 also suggests that the relation between *PEAD* and ρ^2 is steeper when *k* is larger. These considerations lead to our third hypothesis:

Hypothesis 3: *PEAD is positively related to ρ^2 and θ, and the relation between PEAD and ρ^2 is steeper when k is larger.*

4. Data sources, variable measurement, and descriptive statistics

4.1. Data sources

We obtain the data required for our analysis from the following sources: the Center for Research in Security Prices (CRSP), Thomson-Reuters Institutional Holdings (13F), the Institutional Brokers' Estimate System (I/B/E/S), NYSE's Trade and Quote (TAQ), and Standard & Poor's Compustat. Our initial sample includes all Compustat firms that have quarterly earnings announcement dates between 2000 and 2012.[14] To ensure the accuracy of earnings announcement dates, we follow DellaVigna and Pollet (2009) and use the earlier date of two earnings announcement dates reported in Compustat and I/B/E/S (if available). We remove observations if the difference in earnings announcement dates reported in these databases is longer than five calendar days. We keep only those firm-quarters with earnings announcement dates falling within 90 days of the fiscal quarter end. We use only those firms with common shares (share code in 10 or 11) traded in the market and included in CRSP.

4.2. Variable measurement and descriptive statistics

In our model, $\tilde{x}$ is a theoretical construct that captures the *surprise in earnings* implied by the complex sets of information contained in corporate earnings reports and earnings forecast reports produced by various market participants (e.g., analysts). As such, $\tilde{x}$ is not directly observable. In this study, we use the difference between actual earnings and a composite measure of earnings forecast constructed from multiple earnings forecasts as an empirical proxy for $\tilde{x}$. Specifically, we first retrieve all individual analyst forecasts with a

[14] To conduct our empirical analysis with a relatively homogeneous sample, our study period begins in 2000. Regulation Fair Disclosure (Reg FD), introduced by the Securities and Exchange Commission (SEC) in 2000, requires all publicly traded companies to disclose material information to all market participants simultaneously. The regulation prohibits selective disclosure, in which some market participants receive value-relevant information before others. Reg FD fundamentally changed how companies communicate with investors by bringing more transparency and more frequent and timely communications.

forecast period in one or two quarters ahead that are issued within 90 calendar days prior to the earnings announcement date from the I/B/E/S unadjusted detail history file and obtain the median value of these forecasts for each earnings announcement. We then estimate earnings surprise (i.e., our empirical proxy for $\tilde{x}$) using the following formula:

$$UE = \frac{\left(Actual\ EPS - Analyst\ EPS\ Forecast\right)_q}{PRICE_q}; \tag{8}$$

where *UE* denotes earnings surprise, *Actual EPS* is the actual earnings per share obtained from the I/B/E/S unadjusted detail history file, *Analyst EPS Forecast* is the median value of analyst forecasts explained above, and $PRICE_q$ is the closing price at the current fiscal quarter end obtained from Compustat. We exclude observations if the absolute value of the actual *EPS* or the absolute value of the median analyst *EPS* forecast is larger than the stock price. We also delete observations if the stock price is lower than $1.

If all traders were to have an access to the I/B/E/S unadjusted detail history file *and* use it to estimate earnings surprise *UE* in equation (8), *UE* would be a closer empirical proxy for earnings signal $\tilde{y}$ than for earnings signal $\tilde{x}$. However, only a subset of traders were to have an access to and make use of the data, *UE* would be considered a reasonable empirical proxy for $\tilde{x}$. In this sense, our empirical analysis could be considered a joint test of our model and the reasonableness of our empirical proxy for unobservable variable $\tilde{x}$.

We use trade and quote data from the TAQ database to calculate order imbalance (*OI*) at the time of the earnings announcement. We measure *OI* by the mean daily difference between the buyer- and seller-initiated dollar volume (i.e., *Buy$Volume – Sell$Volume*) over the two-day event window [0, +1] minus the mean daily difference between the buyer- and seller-initiated dollar volume over the 40-day pre-event window [−41, −2], scaled by the mean daily dollar volume over the 250-day pre-event period [−251, −2]. That is,

$$OI = \frac{\frac{1}{2}\sum_{d=0}^{1}\left(Buy\$Volume - Sell\$volume\right)_d - \frac{1}{40}\sum_{d=-2}^{-41}\left(Buy\$Volume - Sell\$volume\right)_d}{\frac{1}{250}\sum_{d=-2}^{-251}\left(Buy\$Volume + Sell\$Volume\right)_d}. \quad (9)$$

We follow the Ellis, Michaely, and O'Hara (2000) algorithm to determine whether a buyer or a seller initiates a trade.[15]

Figure 2 shows the daily order imbalance from $d = -30$ to $d = 30$ for firms with positive, zero, and negative *UE*. Prior to the earnings announcement date, the daily order imbalances are all close to zero. For positive *UE* firms, the daily order imbalance increases from $d = 0$ to $d = 1$, with the total (cumulative) increase reaching around 8% (4% each day) over the two-day event window. It resumes to its pre-event level beginning from $d = 2$. For negative *UE* firms, the daily order imbalance decreases from $d = 0$ to $d = 1$, and the largest decrease occurs on $d = 1$. The total decrease over the two-day event window is about 5.2%. This result suggests a positive correlation between order imbalance and earnings surprise (i.e., $\rho > 0$) and is consistent with the notion that the liquidity demander trades on superior earnings information. Unlike positive *UE* firms, there appears to be a post-earnings announcement trading drift for negative *UE* firms. For zero *UE* firms, we find no abnormal order imbalance over the two-day event period.

We define and measure other variables as follows. *AR* is the two-day [0, +1] size and B/M adjusted return, *BHAR* is the 60-day [+2, +61] size and B/M adjusted return, and *U* is the 62-day [0, +61] size and B/M adjusted return.[16] *PI* is the two-day [0, +1] adverse selection

[15] The quoted ask and bid prices are based on National Best Bid and Offer (NBBO) quotes. We follow the NBBO algorithm provided by WRDS to obtain NBBO quotes. Following Bessembinder (2003), we match each trade with contemporaneous quotes.

[16] Most prior studies of the post-earnings announcement drift use either a 60-day window/3-month period (Bartov et al., 2000; Garfinkel and Sokobin, 2006; Ng et al., 2008; Chordia et al., 2009; Hirschleifer, 2009; Chung and Hrazdil, 2011) or a 12-month period (Ng et al., 2008; Atkins et al., 2012). Mendenhall (2004) chooses an interval from one day after the earnings announcement date through the next earnings announcement date (roughly a 3-month period). Ke and Ramalingegowda (2005) measure cumulative abnormal returns over the 3-, 6-, and 9-month period after the earnings announcement. Francis et al. (2007) focus on the 6-month holding period after the earnings announcement.

component/price impact measure from Huang and Stoll (1996). *ESPREAD* is the trade-weighted effective spread during the two-day event period. *PIN* is the probability of informed trading measure from Easley, Kiefer, O'Hara, and Paperman (1996). *OD* is the measure of opinion divergence from Garfinkel and Sokobin (2006). *MVE* is the market value of equity, *NAF* is the number of analysts following the firm, and *NII* is the number of institutional investors that hold the firm's shares. *TURN* is the average daily turnover ratio during the 50-day period prior to earnings announcement. *BETA* and *SIGMA* are systematic and unsystematic risk, respectively. *EAPER* is the measure of earnings persistence and *MBR* is the market-to-book asset ratio. *LOSS* is equal to one if announced earnings are negative and zero otherwise. *EAVOL* is earnings volatility and *PREANN* is the 40-day preannouncement size and B/M adjusted return. Appendix C provides details regarding the construction of these variables. Table 1 reports descriptive statistics.

5. Empirical results

In Section 5.1, we test our hypothesis (Hypothesis 1) regarding how our model-implied information asymmetry measure is related to various information asymmetry proxies. In Section 5.2, we test our hypothesis (Hypothesis 2) regarding whether the degree of market underreaction to earnings announcements is positively related to information asymmetry and ρ^2 and negatively related to the information content of earnings. Section 5.3 tests our hypothesis (Hypothesis 3) regarding whether *PEAD* can be explained by ρ^2, θ, and *k*.

To reduce the effect of outliers, we divide each firm/stock attribute (except *AR*, *BHAR*, and *U*) into 11 groups, assign numeric values −1, −0.8, −0.6, −0.4, −0.2, 0, 0.2, 0.4, 0.6, 0.8, and 1 to these groups, and use these numeric values in empirical analysis. We use raw values of *AR*, *BHAR*, and *U* (instead of their group numeric values) because they are the variables that we attempt to explain. We scale *OI* and *UE* so that their standard deviations are equal to the standard deviation of U. The purpose of the scaling is to ensure that the estimates of k and ρ are within their theoretically permissible range (i.e., between 0 and 1) so that we can obtain a reasonable estimate of $\hat{\theta}$.[17]

5.1. Empirical results for Hypothesis 1

To examine how our model-implied information asymmetry measure is related to a firm/stock attribute, we calculate $\hat{\rho}$ and $\hat{k}$ for each of the 11 groups of the firm/stock attribute. Specifically, we obtain $\hat{k}$ by regressing U on UE and $\hat{\rho}$ by regressing OI on UE using their time-series and cross-sectional observations in each group.[18] We then obtain our model-implied information asymmetry measure (i.e., $\hat{\theta}$) from equation (7) using the values of $\hat{\rho}$ and $\hat{k}$. Finally, we calculate the Pearson and Spearman rank correlation coefficients between the group numeric values (i.e., from −1 to 1) of the firm/stock attribute and the values of $\hat{\rho}$, $\hat{k}$, and $\hat{\theta}$ across the 11 groups. We repeat the above process for each of the seven firm/stock attributes (*PI*, *ESPREAD*, *PIN*, *OD*, *MVE*, *NAF*, and *NII*) that are likely to vary with information asymmetry. It is important to note that we estimate $\hat{\rho}$ and $\hat{k}$ independently of these firm/stock attributes. Hence, a significant correlation between $\hat{\theta}$ and these firm/stock attributes helps us gain confidence that the positive correlation between order imbalance and earnings surprise (shown in Figure 2) is driven by information asymmetry.

Table 2 reports $\hat{\rho}$, $\hat{k}$, and $\hat{\theta}$ as well as the Pearson correlation and Spearman rank

[17] This approach ensures that the estimates do not exceed the theoretical maximum value, but does not guarantee that the estimates will not be negative. In our sample we find that only a few firms have negative $\hat{k}$ or $\hat{\rho}$. To calculate $\hat{\theta}$, we set them to be zero if they are negative.

[18] Note that the regression coefficient estimated using our scaled dependent and independent variables is the correlation coefficient between the two variables. We employ the regression approach, instead of directly calculating the correlation coefficient between *OI* and *UE*, to improve the precision of ρ estimates (see Section 5.2).

correlation between firm/stock attributes and $\hat{\rho}$, $\hat{k}$, and $\hat{\theta}$. Panel A shows that $\hat{\theta}$ increases with *PI* (Pearson correlation = 0.797, p value = 0.003), *ESPREAD* (Pearson correlation = 0.929, p value = 0.000), *PIN* (Pearson correlation = 0.629, p value = 0.038), and *OD* (Pearson correlation = 0.754, p value = 0.007), and decreases with *MVE* (Pearson correlation = −0.924, p value = 0.000), *NAF* (Pearson correlation = −0.971, p value = 0.000), and *NII* (Pearson correlation = −0.928, p value = 0.000). The Spearman rank correlation produces almost identical results, except for *PIN* (Spearman correlation = 0.555, p value = 0.077). These results are consistent with our Hypothesis 1 that the model-implied information asymmetry measure is positively related to the price impact of a trade, the effective spread, the probability of informed trading, and the divergence of investors' opinions, and negatively related to firm size, the number of analysts, and the number of institutional investors.

We find that $\hat{k}$ increases with *ESPREAD*, *PIN*, and *OD*, and decreases with *MVE*, *NAF*, and *NII*. The negative association of $\hat{k}$ with *MVE*, *NAF*, and *NII* is consistent with the finding of prior research (e.g., Atiase, 1985; Shores, 1990; El-Gazzar, 1998) that the information content of earnings is smaller for larger firms and those firms that are followed by more analysts.

5.2. Empirical results for Hypothesis 2

5.2.1. Estimation of $\hat{\rho}$, $\hat{k}$, and $\hat{\theta}$

In this section, we test the hypothesis that the degree of underreaction increases with $\hat{\theta}$ and $\hat{\rho}^2$, but decreases with $\hat{k}$. Instead of estimating $\hat{\rho}$, $\hat{k}$, and $\hat{\theta}$ for each firm/stock attribute group (as in Table 2), we calculate these values for each firm-quarter to fully utilize data and improve the precision of estimates. For this, we first estimate the following regression model:

$$OI_{iq} = \gamma_0 + \rho_{iq} \times UE_{iq} + \varepsilon_{iq}, \tag{10}$$

$$\rho_{iq} = \gamma_1 PI_{iq} + \gamma_2 ESPREAD_{iq} + \gamma_3 OD_{iq} + \gamma_4 MVE_{iq} + \gamma_5 NAF_{iq} + \gamma_6 NII_{iq} + \gamma_7 TURN_{iq}$$
$$+ \gamma_8 BETA_{iq} + \gamma_9 SIGMA_{iq} + \gamma_{10} EAPER_{iq} + \gamma_{11} MBR_{iq} + \gamma_{12} LOSS_{iq} + \gamma_{13} EAVOL_{iq}$$
$$+ \gamma_{14} PREANN_{iq} + \sum_{n=1}^{10} \gamma_{n+14} IndustryDummy_{n,iq};$$

where subscript *iq* indicates firm *i* and quarter *q* of earnings announcement.[19] Because we use the scaled values of OI_{iq} and UE_{iq} (with the same standard deviation as the standard deviation of U), ρ_{iq} is the correlation coefficient between order imbalance (OI_{iq}) and unexpected earnings (UE_{iq}), which is identical to the square root of our measure of underreaction for firm *i* in quarter *q* (i.e., ρ_{iq}^2).

We use regression model (10) to estimate ρ instead of directly calculating the correlation coefficient between order imbalance and earnings surprise for each firm because we have only quarterly observations of these variables, which are not adequate for a reliable estimation of the correlation coefficient between the two variables for each firm. Another important advantage of using regression model (10) over a direct calculation of the correlation coefficient between order imbalance and earnings surprise is that we could increase the precision of ρ estimates by utilizing the firm/stock attributes that are theoretically related to ρ. Equation (5) shows that ρ is related to both information asymmetry (θ) and information content (k). We include various firm/stock attributes that are likely associated with information asymmetry and information content to improve the precision of ρ estimates. Hence, our estimation method is similar, in spirit, to the one used in Pastor and Stambaugh (2003). Pastor and Stambaugh (2003) assume that each stock's liquidity beta is a function of various firm attributes [see their equations (10) and (11)] to increase the precision of beta estimates.

For instance, we include *PI*, *ESPREAD*, *OD*, *MVE*, *NAF*, and *NII* in the regression because they may be associated with the information asymmetry (θ) associated with earnings.[20]

[19] We use the group numeric values of each variable than its raw values in the regression.
[20] We do not include *PIN* in the model because including it results in a significant reduction in sample size.

TURN, a proxy for liquidity, might be related to the liquidity demander's incentive to exploit information advantage. *BETA* controls for systematic risk and *SIGMA* controls for the cost of arbitrage and unsystematic risk. Prior research suggests that the information content of earnings varies with earnings persistence and growth opportunities. *EAPER* controls for earnings persistence and *MBR* controls for growth opportunities. We include *LOSS* because prior research shows that negative earnings have smaller information content (e.g., Hayn, 1995; Basu, 1997).[21] *EAVOL* controls for earnings volatility. High earnings volatility might be associated with low information content. *PREANN* controls for the preannouncement information environment. Finally, to control for industry effects, we include industry dummy variables using the Fama-French 10 industry classification.

Using γ estimates from regression model (10), we then calculate the firm-specific and time-varying estimates of ρ, $\hat{\rho}_{iq}$, using the following equation:

$$\begin{aligned}\hat{\rho}_{iq} &= \hat{\gamma}_1 PI_{iq} + \hat{\gamma}_2 ESPREAD_{iq} + \hat{\gamma}_3 OD_{iq} + \hat{\gamma}_4 MVE_{iq} + \hat{\gamma}_5 NAF_{iq} + \hat{\gamma}_6 NII_{iq} \\ &+ \hat{\gamma}_7 TURN_{iq} + \hat{\gamma}_8 BETA_{iq} + \hat{\gamma}_9 SIGMA_{iq} + \hat{\gamma}_{10} EAPER_{iq} + \hat{\gamma}_{11} MBR_{iq} + \hat{\gamma}_{12} LOSS_{iq} \\ &+ \hat{\gamma}_{13} EAVOL_{iq} + \hat{\gamma}_{14} PREANN_{iq} + \sum_{n=1}^{10} \hat{\gamma}_{n+14} IndustryDummy_{n,iq}.\end{aligned} \tag{11}$$

Note that $\hat{\rho}_{iq}$ captures cross-firm and cross-quarter variations in ρ to the extent that the variations are reflected in the firm/stock attributes used in the estimation. It is also important to note that the above estimation of ρ is not possible without our analytical results.

Similarly, we estimate the following regression model using the 62-day [0, +61] size and B/M adjusted return, *U*, as the dependent variable to calculate the information content of earnings (*k*):[22]

---

[21] Hayn (1995) argues that shareholders have a liquidation option, and therefore returns do not respond as strongly to loss as they do to profit. Basu (1997) argues that due to accounting conservatism, bad news is reflected in earnings in a more timely manner than is good news. *ERC* thus is relatively smaller for losses.

[22] Recall that we defined the information content of earnings by the coefficient on $\tilde{x}$ in the regression of the final payoff ($\tilde{u}$) on earnings surprise ($\tilde{x}$) in Section 2.1.

$$U_{iq} = \beta_0 + k_{iq} \times UE_{iq} + \varepsilon_{iq}, \tag{12}$$

$$k_{iq} = \beta_1 PI_{iq} + \beta_2 ESPREAD_{iq} + \beta_3 OD_{iq} + \beta_4 MVE_{iq} + \beta_5 NAF_{iq} + \beta_6 NII_{iq} + \beta_7 TURN_{iq} + \beta_8 BETA_{iq} + \beta_9 SIGMA_{iq} + \beta_{10} EAPER_{iq} + \beta_{11} MBR_{iq} + \beta_{12} LOSS_{iq} + \beta_{13} EAVOL_{iq} + \beta_{14} PREANN_{iq} + \sum_{n=1}^{10} \beta_{n+14} IndustryDummy_{n,iq} + \varepsilon_{iq}.$$

We include the same set of firm/stock attributes in regression model (12) that we include in regression model (10) to improve the precision of $\hat{k}$ because Table 2 shows that firm/stock attributes that are associated with $\hat{\theta}$ are also associated with $\hat{k}$. Using β estimates from regression model (12), we then calculate $\hat{k}_{iq}$ using the following equation:

$$\hat{k}_{iq} = \hat{\beta}_1 PI_{iq} + \hat{\beta}_2 ESPREAD_{iq} + \hat{\beta}_3 OD_{iq} + \hat{\beta}_4 MVE_{iq} + \hat{\beta}_5 NAF_{iq} + \hat{\beta}_6 NII_{iq} + \hat{\beta}_7 TURN_{iq} + \hat{\beta}_8 BETA_{iq} + \hat{\beta}_9 SIGMA_{iq} + \hat{\beta}_{10} EAPER_{iq} + \hat{\beta}_{11} MBR_{iq} + \hat{\beta}_{12} LOSS_{iq} + \hat{\beta}_{13} EAVOL_{iq} + \hat{\beta}_{14} PREANN_{iq} + \sum_{n=1}^{10} \hat{\beta}_{n+14} IndustryDummy_{n,iq}. \tag{13}$$

Finally, we plug the values of $\hat{\rho}_{iq}$ and $\hat{k}_{iq}$ into equation (7) to calculate $\hat{\theta}_{iq}$. Panel A in Table 3 shows the results of regression models (10) and (12). Consistent with the results in the previous section, we find that in regression model (10), the estimated coefficients on the interaction terms with *PI*, *ESPREAD*, and *OD* are all positive and significant, suggesting a positive relation between these stock attributes and ρ. In contrast, the coefficients on the interaction terms with *NAF* and *NII* are negative and significant, suggesting a negative relation between these attributes and ρ. The information content of earnings (*k*) decreases with firm size (*MVE*), analyst following (*NAF*), turnover rate (*TURN*), and earnings volatility (*EAVOL*). Consistent with the findings of Hayn (1995) and Basu (1997), we also find that the information content is lower for negative earnings (*LOSS*). In contrast, the information content of earnings increases with the divergence of investors' opinions (*OD*) and unsystematic risk (*SIGMA*).Panel B in Table 3 presents the descriptive statistics of $\hat{\rho}_{iq}$, $\hat{k}_{iq}$, and $\hat{\theta}_{iq}$. The mean values of $\hat{\rho}_{iq}$, $\hat{k}_{iq}$, and $\hat{\theta}_{iq}$ are 0.051, 0.136, and 0.011, respectively. The correlation between $\hat{\rho}_{iq}$ and $\hat{k}_{iq}$ is about 0.686, and the correlation between $\hat{\theta}_{iq}$ and $\hat{k}_{iq}$ is about 0.608 (not tabulated). The results show that about 5% of $\hat{\rho}_{iq}$ and $\hat{k}_{iq}$ estimates have negative values. We obtain qualitatively identical results regardless of whether we set these negative values to zero or drop them from the study sample.

5.2.2. Regression results

Before we formally test the positive relation between the observed degree of market underreaction and ρ^2 (Hypothesis 2), we first measure the average degree of underreaction for our sample firms using the following regression model:

$$\begin{aligned} AR_{iq} &= a_1 + (1-b)\cdot \hat{k}_{iq} UE_{iq} + \varepsilon_{iq} \\ BHAR_{iq} &= a_2 + b\cdot \hat{k}_{iq} UE_{iq} + e_{iq}; \end{aligned}, \tag{14}$$

where $\hat{k}_{iq}$ is estimated from equation (13). Note that the regression coefficient on *UE* in the first model is the estimate of *ERC*, i.e., $ERC = (1-b)\cdot \hat{k}_{iq}$ and the regression coefficient on *UE* in the second model is the estimate of *PEAD*, i.e., $PEAD = b\cdot \hat{k}_{iq}$.[23] Hence *b* is our estimate of the degree of underreaction or the proportion of the information content of earnings ($\hat{k}_{iq}$) that fails to get into price at the time of earnings announcements.

We employ the Generalized Method of Moments (GMM) to estimate regression model (14). The weighting matrix is based on the Newey-West weighting scheme with three lags. We use a constant term and $\hat{k}_i \times UE_{iq}$ as the instrumental variables, where $\hat{k}_i$ is the firm-specific

[23] See Section 2.2.

average of $\hat{k}_{iq}$.[24] Panel A in Table 4 shows that the estimate of *b* is 29.5%, indicating that, on average, 70.5% of the information content of earning gets into price during the announcement period and the rest (29.5%) gets into price during the post-earnings announcement period.

We now estimate the following model to formally test Hypothesis 2:

$$\begin{aligned} AR_{iq} &= a_1 + \left[(1-b) - \left(c \cdot \hat{\theta}_{iq} - d \cdot \hat{k}_{iq}\right)\right] \cdot \hat{k}_{iq} UE_{iq} + \varepsilon_{iq} \\ BHAR_{iq} &= a_2 + \left[b + \left(c \cdot \hat{\theta}_{iq} - d \cdot \hat{k}_{iq}\right)\right] \cdot \hat{k}_{iq} UE_{iq} + e_{iq}. \end{aligned}, \tag{15}$$

Note that the regression coefficient on *UE* in the first model is the estimate of *ERC*, i.e., *ERC* $= \left[(1-b) - \left(c \cdot \hat{\theta}_{iq} - d \cdot \hat{k}_{iq}\right)\right] \cdot \hat{k}_{iq}$ and the regression coefficient on *UE* in the second model is the estimate of *PEAD*, i.e., *PEAD* $= \left[b + \left(c \cdot \hat{\theta}_{iq} - d \cdot \hat{k}_{iq}\right)\right] \cdot \hat{k}_{iq}$. Hence, $\left[b + \left(c \cdot \hat{\theta}_{iq} - d \cdot \hat{k}_{iq}\right)\right]$ is our estimate of the degree of underreaction. We use a constant term, $\hat{k}_i \times UE_{iq}$, $\hat{\rho}_i^2 \times \hat{k}_i \times UE_{iq}$, and $\hat{k}_i^2 \times UE_{iq}$ as the instrumental variables, where $\hat{\rho}_i^2$ is the firm-specific average of $\hat{\rho}_{iq}^2$.

The degree of underreaction is hypothesized to be positively associated with θ and negatively associated with *k*. Hence, we expect that both *c* and *d* are positive. Panel B in Table 4 shows that *c* = 10.583 (t-statistic = 16.22) and *d* = 2.456 (t-statistic = 9.44). Note that *c* is larger than *d* and the difference is significant at the 1% level, suggesting that the degree of underreaction is more sensitive to θ than to *k*, which is consistent with our analytical results in Figure 1a and Figure 1b that the degree of underreaction is more strongly related to θ than to *k*.

The regression coefficient *b* in regression model (14) provides an estimate of the average degree of underreaction. However, regression model (14) is unable to provide any insight regarding the degree of underreaction for each firm-quarter. Earlier, we show

[24] We use firm-specific average value because $\hat{k}_{iq}$ contains measurement error, which will enter into the regression error term. Thus, we use firm-specific average (i.e., $\hat{k}_i$) because it is correlated with $\hat{k}_{iq}$ but is not correlated with measurement error.

analytically that the square of the correlation coefficient between order imbalance and the earnings surprise (ρ^2) is a measure of underreaction [i.e., $ERC = k \cdot (1-\rho^2)$ and $PEAD = k \cdot \rho^2$]. To assess the empirical validity of ρ^2 as a measure of underreaction for each firm-quarter, we estimate the following model:

$$\begin{aligned} AR_{iq} &= a_1 + \left[(1-b) - c \cdot \hat{\rho}_{iq}^2\right] \cdot \hat{k}_{iq} UE_{iq} + \varepsilon_{iq} \\ BHAR_{iq} &= a_2 + \left[b + c \cdot \hat{\rho}_{iq}^2\right] \cdot \hat{k}_{iq} UE_{iq} + e_{iq}; \end{aligned}, \tag{16}$$

where $\hat{\rho}_{iq}^2$ is our empirical estimate of ρ^2 for stock i in quarter q calculated from equation (11). Note that the regression coefficient ($b + c \cdot \hat{\rho}_{iq}^2$) on UE_{iq} is the observed value of the degree of underreaction, whereas $\hat{\rho}_{iq}^2$ is the analytically predicted value of the degree of underreaction. We assess the empirical validity of our analytical construct ρ^2 by looking at the sign and statistical significance of the estimated value of *c* in the second model. We use a constant term, $\hat{k}_i \times UE_{iq}$, and $\hat{\rho}_i^2 \times \hat{k}_i \times UE_{iq}$ as the instrumental variables.

Since ρ^2 is hypothesized to be positively associated with the degree of underreaction, we expect *c* to be positive and significant. Panel C in Table 4 shows that *c* is 20.564 (t-statistic = 11.70), which is consistent with our model prediction. We find that the estimate of *b* is 0.125 (t-statistic = 8.13), which implies that even when ρ^2 is zero, there is still about 12.5% of information content that does not get into price at the time of earnings announcement.[25] Thus, about 42.4% (12.5% divided by 29.5%) of the market underreaction cannot be explained by our model.[26]

---

[25] We do not use the estimated *b* in regression model (15) to infer this value because the degree of underreaction is a nonlinear function of θ and *k* and thus the estimated *b* will be biased. Since ρ^2 summarizes the effect of θ and *k* and relates their combined effect to the degree of underreaction in a linear fashion, we expect estimated *b* in regression model (16) to be unbiased.

[26] The results of the over-identifying restrictions test indicate that models (14), (15), and (16) are not misspecified and that the instrumental variables are not correlated with regression error terms.

As a robustness check, we use the three-stage least squares (3SLS) and full-information maximum likelihood (FIML) methods to estimate the above three models and report the results in Panel D, Panel E, and Panel F, respectively. Unlike the GMM and 3SLS methods, FIML does not require instrumental variables. To deal with the potential bias due to the measurement errors associated with $\hat{\rho}_{iq}^2$, $\hat{k}_{iq}$, and $\hat{\theta}_{iq}$, we replace them with $\hat{\rho}_i^2$, $\hat{k}_i$, and $\hat{\theta}_i$ in the above three models, where $\hat{\rho}_i^2$ and $\hat{k}_i$ are firm-specific average values and $\hat{\theta}_i$ is calculated using $\hat{\rho}_i^2$ and $\hat{k}_i$. Thus, the estimation results based on the FIML method should be interpreted as the unconditional effect of θ, k, and ρ^2 on the degree of underreaction.

Panel D, Panel E, and Panel F show that the estimation results based on 3SLS and FIML are very similar to the results based on GMM. Note that the FIML results in Panel D and Panel F suggest that about 52.27% (0.1464 divided by 0.2801) of the market underreaction cannot be explained by the unconditional ρ^2. This figure is larger than 42.4% when the ρ^2 estimate is time-varying (the GMM and 3SLS results). Therefore, conditional ρ^2 seems to explain a greater proportion of the degree of underreaction than unconditional ρ^2.[27]

Overall, empirical results are supportive of Hypothesis 2 that the degree of underreaction increases with information asymmetry and ρ^2 and decreases with the information content of earnings.

5.3. Empirical results for Hypothesis 3

In this section, we test our third hypothesis that *PEAD* is predictable by $\hat{\theta}$, $\hat{k}$, and $\hat{\rho}^2$ and that the relation between *PEAD* and $\hat{\rho}^2$ is steeper when $\hat{k}$ is larger. We perform out-of-

[27] The difference between the conditional and unconditional effect cannot be attributed to the econometric methods that we employed. When we use GMM and 3SLS to estimate the unconditional effect, we find very similar results.

sample tests. The first rolling window starts from the first (calendar) quarter of 2000 to the fourth quarter of 2002 (a 12-quarter estimation period). We estimate equation (10) and equation (12) to obtain $\hat{\gamma}$ and $\hat{\beta}$ using all firms that announce earnings during the 12-quarter period. Next, based on equation (11) and (13), we calculate $\hat{\rho}_{iq}^2$, $\hat{k}_{iq}$, and $\hat{\theta}_{iq}$ using $\hat{\gamma}$, $\hat{\beta}$, and firm/stock attributes of all firms that announce earnings in the first quarter of 2003. This completes the estimation for the first rolling window. We repeat the above procedure to calculate $\hat{\rho}_{iq}^2$, $\hat{k}_{iq}$, and $\hat{\theta}_{iq}$ until the fourth quarter of 2012.

To test whether $\hat{\theta}_{iq}$ and $\hat{\rho}_{iq}^2$ can predict *PEAD*, for each quarter from the first quarter of 2003 to the fourth quarter of 2012, we divide the sample into five groups according to *UE* (raw values of *UE* before conversion to group numeric values). We also divide the sample into five groups according to $\hat{\theta}_{iq}$ and $\hat{\rho}_{iq}^2$, respectively. We then calculate the average *BHAR* for each of the 25 groups. Panel A in Table 5 reports the results when the 25 groups are formed by *UE* and $\hat{\theta}_{iq}$ and Panel B reports the results when the 25 groups are formed by *UE* and $\hat{\rho}_{iq}^2$.

Panel A shows the average *BHAR* for each of the 10 groups in the highest and lowest *UE* quintiles. As $\hat{\theta}_{iq}$ increases, the average *BHAR* for the lowest *UE* firms decreases by 2.04 percentage points (t-statistic = −3.20) from 0.43% to −1.61%. As $\hat{\theta}_{iq}$ increases, however, the average *BHAR* for the highest *UE* firms increases by 1 percentage point (t-statistic = 1.82) from 2.52% to 3.52%. The differences in *BHAR* between the highest and lowest *UE* firms are all positive and significant. As $\hat{\theta}_{iq}$ increases, this difference increases by 3.04 percentage points (t-statistic = 3.62) from 2.09% to 5.13%. These results suggest that $\hat{\theta}_{iq}$ has strong predictive power for *PEAD*. Panel B shows similar results. As $\hat{\rho}_{iq}^2$ increases, the difference in *BHAR* between the highest *UE* and lowest *UE* firms increases by 3.15 percentage points (t-statistic =

3.72).

Panel C reports the subsample results when $\hat{k}_{iq}$ is large and when $\hat{k}_{iq}$ is small, respectively. For each quarter, we divide the sample into five groups according to $\hat{k}_{iq}$. Firms that belong to the first and second quintiles have small $\hat{k}_{iq}$ and firms that belong to the fourth and fifth quintiles have large $\hat{k}_{iq}$. We replicate the results in Panel B using these subsample o+bservations. Panel C shows that when $\hat{k}_{iq}$ is small, as $\hat{\rho}^2_{iq}$ increases, the difference in *BHAR* increases by 1.27 percentage points (t-statistic = 0.65), from 2.70% to 3.97%. When $\hat{k}_{iq}$ is large, however, as $\hat{\rho}^2_{iq}$ increases, the difference in *BHAR* increases by 4.89 percentage points (t-statistic = 3.08) from 0.39% to 5.28%. These results are consistent with Hypothesis 3 that the relation between *PEAD* and ρ^2 is steeper for firms with larger *k*.

6. Predictive power of $\hat{\theta}_{iq}$ and $\hat{\rho}^2_{iq}$ relative to that of other firm/stock attributes

Prior research shows that *PEAD* is smaller for firms with smaller transaction costs (Ng, Rusticus, and Verdi, 2008), larger size (Foster et al., 1984; Bernard and Thomas, 1989), more analyst following (Bhushan, 1994; Gleason and Lee, 2003), and more institutional investors (Bartov et al., 2000; Ke and Ramalingegowda, 2005; Atkins et al., 2012). In this section we evaluate the predictive power of $\hat{\theta}_{iq}$ and $\hat{\rho}^2_{iq}$ relative to that of these firm/stock attributes. Our metric of underreaction is about underreaction arising from information asymmetry; how good it is as a measure of underreaction relative to other metrics is essentially an empirical question. If the predictive power of our metric is greater than that of other firm/stock attributes, the extra predictive power would come from the fact that our metric more closely and directly measures underreaction than these attributes.

We estimate the following regression model to help answer the above question:

$$BHAR_{iq} = \beta_0 + \Psi_{iq} \times UE_{iq} + \sum_{n=1}^{10} \beta_{n+13} IndustryDummy_{n,iq} + \varepsilon_{iq},$$

$$\begin{aligned} \Psi_{iq} = \beta_1 + \beta_2 \hat{\theta}_{iq} + \beta_3 \hat{\rho}_{iq}^2 + \beta_4 \hat{k}_{iq} + \beta_5 \rho_{iq}^2 \hat{k}_{iq} + \beta_6 PI_{iq} + \beta_7 ESPREAD_{iq} + \beta_8 OD_{iq} \\ + \beta_9 MVE_{iq} + \beta_{10} NAF_{iq} + \beta_{11} NII_{iq} + \beta_{12} TURN_{iq} + \beta_{13} PIN_{iq}; \end{aligned} \tag{17}$$

where the regression coefficient on *UE* is the estimate of *PEAD*, i.e., $PEAD_{iq} = \Psi_{iq}$.[28]

Panel A of Table 6 reports the results when we include *UE* and the interaction variable between *UE* and each firm/stock attribute in the regression. Column (2) shows that the coefficient on the interaction variable between *UE* and *ESPREAD* is positive and significant at the 1% level, which is consistent with the finding of Ng, Rusticus, and Verdi (2008) that firms with smaller transaction costs provide lower abnormal returns for the *PEAD* trading strategy. Columns (3) through (6) show that the coefficients on the interaction variable between *UE* and *MVE*, *NAF*, or *NII* are all negative and significant at the 5% level. These results are all consistent with the findings of prior research that *PEAD* is smaller for firms with larger size, more analyst following, and more institutional investors. Column (7) shows that *PEAD* is smaller for firms with higher share turnover ratios, which is consistent with the positive relation between *PEAD* and transaction costs documented in Ng, Rusticus, and Verdi (2008).[29] As expected, column (8) shows that *PEAD* increases with the probability of informed trading (*PIN*). Column (1) shows that although the coefficient on *PI* is positive as expected, it is not statistically significant.

Table 6 Panel B columns (1) through (6) show that the coefficients on both $\hat{\theta}_{iq} \times UE_{iq}$ and $\hat{\rho}_{iq}^2 \times UE_{iq}$ are all positive and significant at the 1% level, regardless of whether firm/stock

[28] See Proposition 2.
[29] This is because transaction costs (*ESPREAD*) decreases with trading volume.

attributes are included or excluded.[30] The results also show that the coefficients on the interaction variables between *UE* and each of two direct measures of information asymmetry (*PI* and *PIN*) are not significant when either $\hat{\theta}_{iq}$ or $\hat{\rho}_{iq}^2$ is included in the model, suggesting that the effects of *PI* and *PIN* are subsumed by the effects of $\hat{\theta}_{iq}$ or $\hat{\rho}_{iq}^2$.

More importantly, we find that the predictive power of $\hat{\theta}_{iq}$ and $\hat{\rho}_{iq}^2$ is higher than that of those firm attributes (i.e., *ESPREAD, OD, MVE, NAF*, and *NII*) that prior studies employ to explain the post-earnings announcement drift. We interpret this result as evidence that our theory-based constructs of information asymmetry and underreaction (i.e., $\hat{\theta}_{iq}$ and $\hat{\rho}_{iq}^2$) provide a more accurate and timely measurement of underreaction (and thus, a more accurate prediction of *PEAD*) than these *ad hoc* measures of the firm's information environment used in prior research.

The last two columns in Table 6 Panel B show the results when we add two interaction variables (i.e., $\hat{k}_{iq} \times UE_{iq}$ and $\hat{\rho}_{iq}^2 \times \hat{k}_{iq} \times UE_{iq}$) in the regression. We find that the coefficients on $\hat{\rho}_{iq}^2 \times \hat{k}_{iq} \times UE_{iq}$ are positive and significant, indicating that the effect of $\hat{\rho}_{iq}^2$ on *PEAD* increases with $\hat{k}_{iq}$, which is consistent with Hypothesis 3. The coefficients on $\hat{\rho}_{iq}^2 \times UE_{iq}$ are not significantly different from zero, indicating that *PEAD* is unrelated to $\hat{\rho}_{iq}^2$ when the information content of earnings (k) is zero. We find that the coefficients on $\hat{k}_{iq} \times UE_{iq}$ are negative and significant, indicating *PEAD* decreases with $\hat{k}_{iq}$ when $\hat{\rho}_{iq}^2$ is zero. According to our model (Proposition 2), *PEAD* decreases with k when $k(2-k\theta)>1$. From Corollary 1 and

[30] Our results complement the findings of Ayers et al. (2011) who report a negative relation between *PEAD* and the interaction term between order imbalance at the time of the earnings announcement and earnings surprise. While we show that higher degree of underreaction leads to larger *PEAD*, Ayers et al. (2011) show that given the same degree of underreaction, stronger trade reaction at the time of earnings announcement is associated with smaller *PEAD*. Chordia, Roll, and Subrahmanyam (2002) and Chordia and Subrahmanyam (2004) also analyze the relation between order imbalance and stock returns.

Proposition 3 and from Figure 1a and Figure 1b, ρ^2 approaches to zero when θ approaches to zero and *k* approaches to 1. But small θ and high *k* also imply that $k(2-k\theta)>1$, which is the case when the relation between *PEAD* and *k* is negative. Thus, the negative coefficient on $\hat{k}_{iq}\times UE_{iq}$ is consistent with our model prediction.

7. Profitability of trading strategies based on $\hat{\rho}^2$ and $\hat{\theta}$

In this section we analyze the profitability of portfolio strategies that are based on $\hat{\rho}^2$ and $\hat{\theta}$. To the extent that the post-earnings announcement drift is predictable by $\hat{\rho}^2$ and $\hat{\theta}$, it may be possible to earn positive abnormal returns using a $\hat{\rho}^2$-based or a $\hat{\theta}$-based trading strategy. Since *PEAD* depends on both ρ^2 and *k*, we expect larger profits from stocks/firms with greater $\hat{k}$.

One possible trading strategy is as follows. The first rolling window starts from January 2000 to December 2002. We estimate $\hat{\rho}_{iq}$, $\hat{k}_{iq}$, and $\hat{\theta}_{iq}$ for all firms that announce earnings during this period using equations (10) through (13) and calculate their mean values ($\hat{\rho}_i$, $\hat{k}_i$, and $\hat{\theta}_i$) for each firm during the period.[31] Then, we divide all firms that announce earnings from October 2002 to December 2002 into five groups according to *UE* (raw values of *UE* before conversion to group numeric values). We also divide these firms into five groups according to $\hat{\theta}_i$, $\hat{\rho}_i^2$, and $\hat{\rho}_i^2\times\hat{k}_i$, respectively.

For each of the five groups formed by $\hat{\theta}_i$, $\hat{\rho}_i^2$, or $\hat{\rho}_i^2\times\hat{k}_i$, we form a hedge portfolio that is long in the highest *UE* quintile firms and short in the lowest *UE* quintile firms and hold

[31] Using sample observation during the portfolio formation period for the first rolling window, we divide all firm attributes into 11 groups and convert them into 11 group values. We also scale *OI* and *UE* so that they have the same standard deviation as the standard deviation of *U* calculated during this period only.

each of the five hedge portfolios for the month of January 2003. This completes the first rolling window. We repeat the above procedure and calculate the return of the five hedge portfolios for the month of February 2003. The last portfolio holding month is December 2012. Using the monthly returns of the five hedge portfolios, we calculate the Pastor-Stambaugh four-factor alpha for each hedge portfolio. Table 7 reports the annualized alpha of each of the five hedge portfolios. Panel A shows that as $\hat{\theta}_i$ increases, alpha increases from 0.52% to 8.34%. The difference between the two alphas is 7.82% (t-statistic = 2.23). Panel B shows that as $\hat{\rho}_i^2$ increases, alpha increases from 0.38% to 8.62%. The difference between the two alphas is 8.25% (t-statistic = 2.33). Finally, Panel C shows that as $\hat{\rho}_i^2 \times \hat{k}_i$ increases, alpha increases from −3.19% to 8.44%. The difference between the two alphas is 11.64% (t-statistic = 3.52). Thus, these results indicate that trading on $\hat{\theta}_i$ and $\hat{\rho}_i^2$ are both profitable, and trading on additional information on $\hat{k}_i$ will further improve the profit by about 3.39% (i.e., 11.64% − 8.25%). Overall, these results underscore the potential practical value of our empirical metrics of underreaction and information asymmetry for investors and traders.

8. Summary and concluding remarks

In this study we develop an analytical framework that enables us to disentangle the effects of the information content of earnings announcements and the information asymmetry associated with earnings on market underreaction to earnings announcements. Despite the fact that our analytical model is stylized and simple, we believe that our model and empirical proxies are sufficiently intuitive to be successfully applied empirically.

Our analytical model predicts that the degree of market underreaction to earnings announcements and *PEAD* increase with information asymmetry among traders. We provide robust empirical evidence that is consistent with these predictions. We develop a metric for information asymmetry that is defined by two empirically observable variables (i.e., the information content of earnings and the correlation coefficient between earnings surprise and order imbalance), and we show that this metric is highly correlated with various firm/stock attributes that prior research has shown to be good proxies for information asymmetry in the predicted manner. We also show that our theory-based constructs of information asymmetry and underreaction provide a more accurate and timely measurement of underreaction and thus a more timely prediction of *PEAD* than firm/stock attributes that prior studies employ to explain the post-earnings announcement drift. Furthermore, we analyze the profitability of trading strategies based on the empirical metrics developed in our paper and show that these strategies could generate statistically significant excess returns. On the whole, both our analytical results and empirical findings provide strong support for our conjecture that the market reactions to earnings announcements depend not only on the information content of earnings announcements but also on the information asymmetry associated with earnings.

References

Atiase, R. 1985. Predisclosure information, firm capitalization, and security price behavior around earnings announcements. *Journal of Accounting Research* 23: 21-36.

Atkins, B., J. Ng, and R. Verdi. 2012. Investor competition over information and the pricing of information asymmetry. *Accounting Review* 87 (1): 35-58.

Ayers, B. C., O. Z. Li, and P. E. Yeung. 2011. Investor trading and the post-earnings-announcement drift. *Accounting Review* 86 (2): 385-416.

Barberis N., and W. Xiong. 2009. What drives the disposition effect? An analysis of a long-standing preference-based explanation. *Journal of Finance* 64 (2): 751-784

Bartov, E., S. Radhakrishnan, and I. Krinsky. 2000. Investor sophistication and patterns in stock returns after earnings announcements. *Accounting Review* 75(January): 43-63.

Basu, S. 1997. The conservatism principle and the asymmetric timeliness of earnings. *Journal of Accounting and Economics* 24: 3-37.

Battalio, R. H., and R. R. Mendenhall. 2005. Earnings expectations, investor trade size, and anomalous returns around earnings announcements. *Journal of Financial Economics* 77: 289-319.

Bernard, V., and J. Thomas. 1989. Post-earnings announcement drift: Delayed price response or risk premium. *Journal of Accounting Research* 27 (Supplement): 1-36.

Bernard, V., and J. Thomas. 1990. Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics* 13: 305-340.

Bessembinder, H. 2003. Issues in assessing trade execution costs. *Journal of Financial Markets* 6: 233-257.

Bhat, G., and S. Jayaraman. 2009. Information asymmetry around earnings announcements during the financial crisis. Working paper. Olin School of Business. Washington University in St. Louis.

Bhattacharya, N. 2001. Investors' trade size and trading responses around earnings announcements: An empirical investigation. *Accounting Review* 76 (2): 221-244.

Bhushan, R. 1994. An informational efficiency perspective on the post-earnings announcement drift. *Journal of Accounting and Economics* 18: 45-65.

Brennan, M., and A. Subrahmanyam. 1995. Investment analysis and price formation in securities markets. *Journal of Financial Economics* 38: 361-381.

Chordia, R., A. Goyal, G. Sadka, R. Sadka, and L. Shivakumar. 2009. Liquidity and post-earnings announcement drift. *Financial Analyst Journal* 65 (4): 18-33.

Chordia, T., R. Roll, and A. Subrahmanyam. 2002. Order imbalance, liquidity, and market returns. *Journal of Financial Economics* 65 (1): 111-130.

Chordia, T., and A. Subrahmanyam. 2004. Order imbalance and individual stock returns: Theory and evidence. *Journal of Financial Economics* 72 (3): 485-518.

Chung, D., and K. Hrazdil. 2011. Market efficiency and the post-earnings announcement drift. *Contemporary Accounting Research* 28 (3): 926-956.

Coller, M., and T. Yohn. 1997. Management forecasts and information asymmetry: An examination of bid-ask spreads. *Journal of Accounting Research* 35: 181-191.

Collins, D., and S. P. Kothari. 1989. An analysis of intertemporal and cross-sectional determinants of earnings response coefficients. *Journal of Accounting and Economics* 11: 143-181.

DellaVigna, S., and J. Pollet. 2009. Investor inattention and Friday earnings announcements. *Journal of Finance* 64 (2): 709-749.

Dontoh, A., J. Ronen, and B. Sarath. 2003. On the rationality of the post-announcement drift. *Review of Accounting Studies* 8: 69–104.

Easley, D., N. Kiefer, M. O'Hara, and J. Paperman. 1996. Liquidity, information, and infrequently traded stocks. *Journal of Finance* 51: 1405-1436.

El-Gazzar, S. 1998. Predisclosure information and institutional ownership: A cross-sectional examination of market revaluations during earnings announcement periods. *Accounting Review* 73: 119-129.

Ellis, K., R. Michaely, and M. O'Hara. 2000. The accuracy of trade classification rules: Evidence from NASDAQ. *Journal of Financial and Quantitative Analysis* 35 (4): 529-551.

Foster, G., C. Olsen, and T. Shevlin. 1984. Earnings releases, anomalies, and the behavior of security returns. *Accounting Review* 59: 574-603.

Francis, J., R. Lafond, P. Olsson, and K. Schipper. 2007. Information uncertainty and post-earnings-announcement-drift. *Journal of Business Finance & Accounting* 34 (3): 403-433.

Frazzini, A. 2006. The disposition effect and underreaction to news. *Journal of Finance* 61 (4): 2017–2046.

Garfinkel, J., and J. Sokobin. 2006. Volume, opinion divergence, and returns: a study of post-earnings announcement drift. *Journal of Accounting Research* 44: 85-112.

Grossman, S., and J. Stiglitz. 1980. On the impossibility of informationally efficient markets, *American Economic Review* 70: 393-408.

Gleason, C. A., and C. M. Lee. 2003. Analyst forecast revisions and market price discovery. *Accounting Review* 78: 193-225.

Hayn, C. 1995. The information content of losses. *Journal of Accounting and Economics* 20: 125-153.

Hirshleiffer, D., S. Lim, and S. Teoh. 2009. Driven to distraction: extraneous events and underreaction to earnings news. *Journal of Finance* 64 (5): 2289-2235.

Hirshleiffer, D., S. Lim, and S. Teoh. 2011. Limited investor attention and stock market misreactions to accounting information. *Review of Asset Pricing Studies* 1 (1): 35-73.

Hirshleiffer, D., and S. Teoh. 2003. Limited attention, information disclosure, and financial reporting. *Journal of Accounting and Economics* 36: 337-386.

Huang, R., and H. Stoll. 1996. Dealer versus auction markets: A paired comparison of execution costs on NASDAQ and the NYSE. *Journal of Financial Economics* 41: 313-357.

Kaniel, R., S. Liu, G. Saar, and S. Titman. 2012. Individual investor trading and return patterns around earnings announcements. *Journal of Finance* 67 (2): 639-680.

Ke, B., and S. Ramalingegowda. 2005. Do institutional investors exploit the post-earnings announcement drift? *Journal of Accounting and Economics* 39: 25-53.

Kim, O., and R. Verrecchia. 1994. Market liquidity and volume around earnings announcements. *Journal of Accounting and Economics* 17: 41-67.

Kothari, S. P. 2001. Capital market research in accounting. *Journal of Accounting and Economics* 31: 105–231.

Kyle, A. 1985. Continuous auctions and insider trading. *Econometrica* 53 (6): 1315-1335.

Lee, C. 1992. Earnings news and small traders. *Journal of Accounting and Economics* 15: 265-302.

Lee, C., B. Mucklow, and M. Ready. 1993. Spreads, depths, and the impact of earnings information: An intraday analysis. *Review of Financial Studies* 6: 345-374.

Mendenhall, R. 2004. Arbitrage risk and post-earnings-announcement drift. *Journal of Business* 77: 875-894.

Mian, M., and S. Sankaraguruswamy. 2012. Investor sentiment and stock market response to earnings news. *Accounting Review* 87 (4): 1357-1384.

Ng, J., T. Rusticus, and R. Verdi. 2008. Implications of transaction costs for the post-earnings announcement drift. *Journal of Accounting Research* 46 (3): 661-696.

Odean T., 1998. Are investors reluctant to realize their losses? *Journal of Finance* 53 (5):1775-1798.

Pastor, L., and R. F. Stambaugh. 2003. Liquidity risk and expected stock returns. *Journal of Political Economy* 111 (3): 642-685.

Sadka, R. 2006. Momentum and post-earnings-announcement drift anomalies: the role of liquidity risk. *Journal of Financial Economics* 80: 309-349.

Shefrin H., and M. Statman. 1985. The disposition to sell winners too early and ride losers too long: theory and evidence. *Journal of Finance* 40 (3): 777-790

Shleifer, A, and R. Vishny. 1997. The limits of arbitrage. *Journal of Finance* 52: 1540-6261.

Shores, D. 1990. The association between interim information and security returns surrounding earnings announcements. *Journal of Accounting Research* 28: 164-181.

Vega, C. 2006. Stock price reaction to public and private information. *Journal of Financial Economics* 82 (1): 103-133.

Yohn, T. 1998. Information asymmetry around earnings announcements. *Review of Quantitative Finance and Accounting* 11: 165-82.

Zhang, Q., C. Cai, and K. Keasey. 2013. Market reaction to earnings news: A unified test of information risk and transaction costs. *Journal of Accounting and Economics* 56: 251-266.

Zhang, X. 2006. Information uncertainty and stock returns. *Journal of Finance* 61: 105-136.

Table 1
Descriptive statistics

OI is order imbalance at the time of earnings announcement, *UE* is earnings surprise, *AR* is the two-day [0, +1] size and B/M adjusted return, *BHAR* is the 60-day [+2, +61] size and B/M adjusted return, *U* is the 62-day [0, +61] size and B/M adjusted return, *PI* is the two-day [0, +1] adverse selection component/price impact measure, *ESPREAD* is the trade-weighted effective spread during the two-day event period, *PIN* is the probability of informed trading, *OD* is the measure of opinion divergence, *MVE* is the market value of equity, *NAF* is the number of analysts following the firm, *NII* is the number of institutional investors that hold the firm's shares, *TURN* is the average daily turnover ratio during the 50-day period prior to earnings announcement, *BETA* and *SIGMA* are systematic and unsystematic risk, respectively, *EAPER* is the measure of earnings persistence, *MBR* is the market-to-book asset ratio, *LOSS* is equal to one if announced earnings are negative and zero otherwise, *EAVOL* is earnings volatility, and *PREANN* is the 40-day preannouncement size and B/M adjusted return. Appendix C provides details regarding the construction of these variables.

			Standard	Percentile				
Variable	N	Mean	deviation	5th	25th	50th	75th	95th
OI	142,052	1.308	63.489	−41.693	−8.817	0.649	11.493	49.171
UE	142,223	−0.108	3.238	−1.519	−0.070	0.047	0.230	1.266
AR (%)	137,696	0.072	8.855	−13.538	−3.761	0.013	3.968	13.588
BHAR (%)	137,695	0.243	22.571	−31.778	−10.606	−0.757	9.320	33.780
U (%)	137,701	0.417	24.682	−34.964	−11.855	−0.776	10.612	37.969
PI	141,774	0.169	2.535	−1.141	−0.006	0.068	0.256	1.720
ESPREAD	142,223	0.394	0.614	0.030	0.075	0.166	0.423	1.577
PIN	102,865	0.180	0.100	0.070	0.114	0.158	0.219	0.372
OD	141,938	0.932	2.444	−0.399	−0.030	0.295	1.089	4.169
MVE ($million)	142,214	4.796	19.208	0.059	0.240	0.723	2.453	18.298
NAF	142,223	9.187	7.521	1	4	7	13	25
NII	141,824	161.305	179.809	16	57	109	193	490
TURN	141,938	0.912	1.000	0.112	0.349	0.647	1.139	2.555
BETA	141,873	1.074	0.595	0.180	0.661	1.032	1.429	2.110
SIGMA (%)	141,873	2.912	1.749	1.027	1.711	2.464	3.620	6.367
EAPER	139,033	0.298	2.002	−0.140	0.080	0.284	0.509	0.767
MBR	141,182	2.143	2.663	0.878	1.077	1.441	2.304	5.414
LOSS	139,033	0.256	0.437	0	0	0	1	1
EAVOL	137,489	4.242	7.239	0.091	0.648	1.656	4.506	17.029
PREANN (%)	137,701	0.282	20.107	−26.769	-8.837	-0.688	7.620	28.381

Table 2
Correlation between ρ, k, θ and firm/stock attributes

We divide each firm/stock attribute into 11 groups, assign numeric values −1, −0.8, −0.6, −0.4, −0.2, 0, 0.2, 0.4, 0.6, 0.8, and 1 to these groups, and use these numeric values in empirical analysis. We scale *OI* and *UE* so that their standard deviations are equal to the standard deviation of *U*, where *OI* is order imbalance at the time of earnings announcement, *UE* is earnings surprise, and *U* is the 62-day [0, +61] size and B/M adjusted return. To examine how our model-implied information asymmetry measure is related to a firm/stock attribute, we calculate $\hat{\rho}$ and $\hat{k}$ for each of the 11 groups of the firm/stock attribute. Specifically, we obtain $\hat{k}$ by regressing *U* on *UE* and $\hat{\rho}$ by regressing *OI* on *UE* using their time-series and cross-sectional observations in each group. We then obtain our model-implied information asymmetry measure (i.e., $\hat{\theta}$) using the values of $\hat{\rho}$ and $\hat{k}$. Finally, we calculate the Pearson and Spearman rank correlation coefficients between the group numeric values (i.e., from −1 to 1) of the firm/stock attribute and the values of $\hat{\rho}$, $\hat{k}$, and $\hat{\theta}$ across the 11 groups. We repeat the above process for each of the seven firm/stock attributes (*PI*, *ESPREAD*, *PIN*, *OD*, *MVE*, *NAF*, and *NII*), where *PI* is the two-day [0, +1] adverse selection component/price impact measure, *ESPREAD* is the trade-weighted effective spread during the two-day event period, *PIN* is the probability of informed trading, *OD* is the measure of opinion divergence, *MVE* is the market value of equity, *NAF* is the number of analysts following the firm, *NII* is the number of institutional investors that hold the firm's shares. Appendix C provides details regarding the construction of these variables.

Panel A: Portfolio formed by PI

PI	Mean *PI*	$\hat{\rho}$	$\hat{k}$	$\hat{\theta}$
−1	−3.458	0.047	0.146	0.005
−0.8	−0.136	0.057	0.158	0.008
−0.6	−0.017	0.030	0.132	0.002
−0.4	0.013	0.023	0.125	0.001
−0.2	0.037	0.055	0.122	0.007
0	0.069	0.032	0.138	0.002
0.2	0.113	0.052	0.146	0.006
0.4	0.180	0.066	0.143	0.010
0.6	0.296	0.076	0.172	0.014
0.8	0.569	0.085	0.163	0.017
1	4.193	0.095	0.180	0.022
Pearson correlation		0.754	0.596	0.797
		(0.007)	(0.053)	(0.003)
Spearman correlation		0.736	0.564	0.736
		(0.009)	(0.071)	(0.009)

Panel B: Portfolio formed by ESPREAD

ESPREAD	Mean *ESPREAD*	$\hat{\rho}$	$\hat{k}$	$\hat{\theta}$
−1	0.025	0.006	0.095	0.000
−0.8	0.049	0.003	0.118	0.000
−0.6	0.069	0.027	0.127	0.002
−0.4	0.093	0.046	0.141	0.005
−0.2	0.124	0.064	0.135	0.009
0	0.167	0.059	0.150	0.008
0.2	0.228	0.068	0.143	0.011
0.4	0.320	0.058	0.166	0.008
0.6	0.475	0.063	0.177	0.010
0.8	0.783	0.079	0.183	0.015
1	2.000	0.094	0.157	0.021
Pearson correlation		0.923	0.906	0.929
		(0.000)	(0.000)	(0.000)
Spearman correlation		0.873	0.927	0.909
		(0.001)	(0.000)	(0.000)

Panel C: Portfolio formed by PIN				
PIN	Mean *PIN*	$\hat{\rho}$	$\hat{k}$	$\hat{\theta}$
−1	0.064	0.015	0.125	0.000
−0.8	0.093	0.068	0.159	0.011
−0.6	0.110	0.066	0.142	0.010
−0.4	0.125	0.073	0.141	0.012
−0.2	0.141	0.076	0.146	0.013
0	0.158	0.082	0.165	0.016
0.2	0.176	0.057	0.163	0.008
0.4	0.198	0.064	0.157	0.010
0.6	0.228	0.083	0.182	0.017
0.8	0.277	0.079	0.194	0.015
1	0.414	0.079	0.150	0.015
Pearson correlation		0.594	0.677	0.629
		(0.054)	(0.022)	(0.038)
Spearman correlation		0.564	0.627	0.555
		(0.071)	(0.039)	(0.077)

Panel D: Portfolio formed by OD				
OD	Mean *OD*	$\hat{\rho}$	$\hat{k}$	$\hat{\theta}$
−1	−0.634	0.015	0.054	0.000
−0.8	−0.169	0.043	0.081	0.004
−0.6	−0.054	0.059	0.100	0.008
−0.4	0.043	0.049	0.091	0.005
−0.2	0.154	0.067	0.134	0.010
0	0.297	0.066	0.136	0.010
0.2	0.494	0.072	0.144	0.012
0.4	0.781	0.072	0.155	0.012
0.6	1.234	0.056	0.191	0.008
0.8	2.074	0.058	0.218	0.009
1	6.032	0.065	0.344	0.013
Pearson correlation		0.646	0.906	0.754
		(0.032)	(0.000)	(0.007)
Spearman correlation		0.473	0.991	0.764
		(0.142)	(0.000)	(0.006)

Panel E: Portfolio formed by MVE				
MVE	Mean *MVE*	$\hat{\rho}$	$\hat{k}$	$\hat{\theta}$
−1	0.055	0.093	0.186	0.021
−0.8	0.128	0.069	0.173	0.011
−0.6	0.215	0.072	0.158	0.012
−0.4	0.331	0.083	0.162	0.017
−0.2	0.493	0.057	0.156	0.008
0	0.728	0.060	0.136	0.008
0.2	1.093	0.056	0.142	0.007
0.4	1.697	0.039	0.142	0.003
0.6	2.872	0.012	0.102	0.000
0.8	6.080	0.003	0.101	0.000
1	39.061	0.015	0.084	0.000
Pearson correlation		−0.930	−0.954	−0.924
		(0.000)	(0.000)	(0.000)
Spearman correlation		−0.927	−0.964	−0.927
		(0.000)	(0.000)	(0.000)

Panel F: Portfolio formed by NAF				
NAF	Mean *NAF*	$\hat{\rho}$	$\hat{k}$	$\hat{\theta}$
−1	1.000	0.080	0.178	0.016
−0.8	2.000	0.087	0.181	0.019
−0.6	3.000	0.077	0.160	0.014
−0.4	4.481	0.069	0.158	0.011
−0.2	6.000	0.065	0.166	0.010
0	7.000	0.061	0.132	0.008
0.2	8.462	0.038	0.144	0.003
0.4	10.475	0.044	0.115	0.004
0.6	13.358	0.036	0.116	0.003
0.8	17.760	0.024	0.122	0.001
1	27.027	−0.005	0.122	0.000
Pearson correlation		−0.953	−0.917	−0.971
		(0.000)	(0.000)	(0.000)
Spearman correlation		−0.982	−0.873	−0.982
		(0.000)	(0.001)	(0.000)

Panel G: Portfolio formed by NII				
NII	Mean *NII*	$\hat{\rho}$	$\hat{k}$	$\hat{\theta}$
−1	14.887	0.093	0.175	0.021
−0.8	34.133	0.077	0.149	0.014
−0.6	53.000	0.059	0.188	0.009
−0.4	71.564	0.079	0.160	0.015
−0.2	89.348	0.073	0.151	0.012
0	108.691	0.060	0.142	0.008
0.2	131.505	0.046	0.155	0.005
0.4	161.171	0.035	0.136	0.003
0.6	207.666	0.026	0.112	0.002
0.8	292.369	−0.008	0.113	0.000
1	610.217	0.016	0.093	0.001
Pearson correlation		−0.921	−0.879	−0.928
		(0.000)	(0.000)	(0.000)
Spearman correlation		−0.918	−0.845	−0.945
		(0.000)	(0.001)	(0.000)

Table 3
Estimation of ρ, *k*, and θ

Panel A shows how the correlation coefficient between order imbalance and earnings surprise (ρ) and the information content of earnings (k) are related to various firm/ stock attributes. Panel B shows the descriptive statistics of the correlation coefficient between order imbalance and earnings surprise ($\hat{\rho}_{iq}$), the information content of earnings ($\hat{k}_{iq}$), and the model implied information asymmetry measure ($\hat{\theta}_{iq}$). *OI* is order imbalance at the time of earnings announcement, *UE* is earnings surprise, *U* is the 62-day [0, +61] size and B/M adjusted return, *PI* is the two-day [0, +1] adverse selection component/price impact measure, *ESPREAD* is the trade-weighted effective spread during the two-day event period, *OD* is the measure of opinion divergence, *MVE* is the market value of equity, *NAF* is the number of analysts following the firm, *NII* is the number of institutional investors that hold the firm's shares, *TURN* is the average daily turnover ratio during the 50-day period prior to earnings announcement, *BETA* and *SIGMA* are systematic and unsystematic risk, respectively, *EAPER* is the measure of earnings persistence, *MBR* is the market-to-book asset ratio, *LOSS* is equal to one if announced earnings are negative and zero otherwise, *EAVOL* is earnings volatility, and *PREANN* is the 40-day preannouncement size and B/M adjusted return. Appendix C provides details regarding the construction of these variables. Standard errors are adjusted for clustering by firm and time. Figures in parentheses are *t*-statistics.

Panel A: ρ and k regressions

Dependent variables	Variables interacting with *UE*													
	PI	*ESPREAD*	*OD*	*MVE*	*NAF*	*NII*	*TURN*	*BETA*	*SIGMA*	*EAPER*	*MBR*	*LOSS*	*EAVOL*	*PREANN*
OI	1.83***	2.31***	4.65***	0.13	−1.84***	−3.12**	−0.51	−0.96*	−0.64	−0.28	0.78	−0.15	−1.87***	1.33***
	(3.89)	(2.89)	(9.45)	(0.11)	(−2.70)	(−2.26)	(−0.82)	(−1.88)	(−0.88)	(−0.62)	(1.45)	(−0.21)	(−2.96)	(3.18)
U	0.84	1.14	14.16***	−3.59**	−3.33***	−0.51	−2.96***	−0.14	2.18***	0.39	−0.23	−3.97***	−1.49**	0.06
	(1.48)	(1.33)	(21.57)	(−2.21)	(−3.78)	(−0.29)	(−3.75)	(−0.23)	(2.99)	(0.67)	(−0.34)	(−5.01)	(−2.00)	(0.11)

Panel B: Descriptive statistics

Variable	N	Mean	Standard deviation	Min	1st	5th	25th	50th	75th	95th	99th	Max
$\hat{\rho}_{iq}$	135,486	0.051	0.045	−0.126	−0.049	−0.021	0.019	0.050	0.082	0.128	0.157	0.224
$\hat{k}_{iq}$	135,486	0.136	0.092	−0.191	−0.081	−0.023	0.074	0.140	0.201	0.279	0.330	0.444
$\hat{\theta}_{iq}$	135,486	0.011	0.015	0.000	0.000	0.000	0.001	0.006	0.016	0.042	0.067	0.157

***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

Table 4
Testing how underreaction is related to information asymmetry, information content, and ρ^2

Panel A shows the results of the following regression model:

$$AR_{iq} = a_1 + (1-b)\cdot \hat{k}_{iq}UE_{iq} + \varepsilon_{iq},$$
$$BHAR_{iq} = a_2 + b\cdot \hat{k}_{iq}UE_{iq} + e_{iq};$$

where AR_{iq} is the two-day [0, +1] size and B/M adjusted return, $BHAR_{iq}$ is the 60-day [+2, +61] size and B/M adjusted return, UE_{iq} is earnings surprise, and $\hat{k}_{iq}$ is the estimate of the information content of earnings for firm i in quarter q. We employ Generalized Method of Moments (GMM) to estimate the model. The weighting matrix is based on the Newey-West weighting scheme with three lags. We use a constant term and $\hat{k}_i \times UE_{iq}$ as the instrumental variables, where $\hat{k}_i$ is the firm-specific average of $\hat{k}_{iq}$. Panel B shows the results of the following regression model:

$$AR_{iq} = a_1 + \left[(1-b) - \left(c\cdot\hat{\theta}_{iq} - d\cdot\hat{k}_{iq}\right)\right]\cdot \hat{k}_{iq}UE_{iq} + \varepsilon_{iq},$$
$$BHAR_{iq} = a_2 + \left[b + \left(c\cdot\hat{\theta}_{iq} - d\cdot\hat{k}_{iq}\right)\right]\cdot \hat{k}_{iq}UE_{iq} + e_{iq}.$$

We employ Generalized Method of Moments (GMM) to estimate the model. The weighting matrix is based on the Newey-West weighting scheme with three lags. We use a constant term, $\hat{k}_i \times UE_{iq}$, $\hat{\rho}_i^2 \times \hat{k}_i \times UE_{iq}$, and $\hat{k}_i^2 \times UE_{iq}$ as the instrumental variables, where $\hat{\rho}_i^2$ is the firm-specific average of $\hat{\rho}_{iq}^2$. Panel C shows the results of the following regression model:

$$AR_{iq} = a_1 + \left[(1-b) - c\cdot\hat{\rho}_{iq}^2\right]\cdot \hat{k}_{iq}UE_{iq} + \varepsilon_{iq},$$
$$BHAR_{iq} = a_2 + \left[b + c\cdot\hat{\rho}_{iq}^2\right]\cdot \hat{k}_{iq}UE_{iq} + e_{iq};$$

where $\hat{\rho}_{iq}^2$ is our empirical estimate of ρ^2 for stock i in quarter q. We employ Generalized Method of Moments (GMM) to estimate the model. We use a constant term, $\hat{k}_i \times UE_{iq}$, and $\hat{\rho}_i^2 \times \hat{k}_i \times UE_{iq}$ as the instrumental variables. As a robustness check, we use the three-stage least squares (3SLS) and full-information maximum likelihood (FIML) method to estimate the above three models and report the results in Panel D, Panel E, and Panel F, respectively. Unlike the GMM and 3SLS methods, FIML does not require instrumental variables. To deal with the potential bias due to the measurement errors associated with $\hat{\rho}_{iq}^2$, $\hat{k}_{iq}$, and $\hat{\theta}_{iq}$, we replace them with $\hat{\rho}_i^2$, $\hat{k}_i$, and $\hat{\theta}_i$ in the above three models, where $\hat{\rho}_i^2$ and $\hat{k}_i$ are firm-specific average values and $\hat{\theta}_i$ is calculated using $\hat{\rho}_i^2$ and $\hat{k}_i$. Thus, the estimation results based on FIML method should be interpreted as the unconditional effect of θ, k, and ρ^2 on the degree of underreaction. Figures in parentheses are t-statistics.

Panel A: Results of regression model (14) using GMM						
a_1	a_2	b			χ^2	N
0.000*	0.003***	0.295***			0.3750	135,479
(1.75)	(4.67)	(41.71)			[0.9454]	
Panel B: Results of regression model (15) using GMM						
a_1	a_2	b	c	d	χ^2	N
−0.000	0.003***	0.593***	10.583***	2.456***	7.7995	135,479
(−0.03)	(5.27)	(11.55)	(16.22)	(9.44)	[0.1676]	
Panel C: Results of regression model (16) using GMM						
a_1	a_2	b	c		χ^2	N
0.000	0.003***	0.125***	20.564***		0.8575	135,479
(1.30)	(4.84)	(8.13)	(11.70)		[0.9306]	

Panel D: Results of regression model (14) using 3SLS and FIML					
	a_1	a_2	b		
3SLS	0.000*	0.003***	0.295***		
	(1.79)	(4.71)	(54.28)		
FIML	0.001***	0.003***	0.280***		
	(4.26)	(4.51)	(63.30)		
Panel E: Results of regression model (15) using 3SLS and FIML					
	a_1	a_2	b	c	d
3SLS	−0.000	0.003***	0.604***	10.529***	2.513***
	(−0.08)	(5.40)	(16.45)	(22.36)	(13.30)
FIML	0.001***	0.003***	0.418***	11.57***	1.70***
	(2.91)	(5.01)	(21.09)	(28.94)	(14.02)
Panel F: Results of regression model (16) using 3SLS and FIML					
	a_1	a_2	b	c	
3SLS	0.000	0.003***	0.127***	20.265***	
	(1.34)	(4.88)	(11.85)	(18.22)	
FIML	0.001***	0.003***	0.146***	23.287***	
	(3.43)	(4.82)	(22.03)	(29.40)	

***Significant at the 1% level. *Significant at the 10% level.

Table 5
Post-earnings announcement drift of portfolios formed by *UE*, $\hat{\theta}$, $\hat{\rho}^2$, and $\hat{k}$

This table shows whether *PEAD* is predictable by $\hat{\theta}$, $\hat{k}$, and $\hat{\rho}^2$ and the relation between *PEAD* and $\hat{\rho}^2$ is steeper when $\hat{k}$ is larger. We perform out-of-sample tests. The first rolling window starts from the first (calendar) quarter of 2000 to the fourth quarter of 2002 (a 12-quarter estimation period). We estimate $\hat{\gamma}$ and $\hat{\beta}$ using all firms that announce earnings during the 12-quarter period. Next, we calculate $\hat{\rho}_{iq}^2$, $\hat{k}_{iq}$, and $\hat{\theta}_{iq}$ using $\hat{\gamma}$, $\hat{\beta}$, and firm/stock attributes of all firms that announce earnings in the first quarter of 2003. This completes the estimation for the first rolling window. We repeat the above procedure to calculate $\hat{\rho}_{iq}^2$, $\hat{k}_{iq}$, and $\hat{\theta}_{iq}$ until the fourth quarter of 2012. To test whether $\hat{\theta}_{iq}$ and $\hat{\rho}_{iq}^2$ can predict *PEAD*, for each quarter from the first quarter of 2003 to the fourth quarter of 2012, we divide the sample into five groups according to *UE* (raw values before conversion to 11 numeric values). For each quarter, we also divide the sample into five groups according to $\hat{\theta}_{iq}$ and $\hat{\rho}_{iq}^2$, respectively. We then calculate the average *BHAR* for each of the 25 groups. Panel A in Table 5 reports the results when the 25 groups are formed by *UE* and $\hat{\theta}_{iq}$ and Panel B reports the results when the 25 groups are formed by *UE* and $\hat{\rho}_{iq}^2$. Figures in parentheses are *t*-statistics.

Panel A: Portfolios formed by UE and $\hat{\theta}$

UE Quintile	$\hat{\theta}$ *Quintile*					
	1	2	3	4	5	5 − 1
1	0.43%	−0.53%	−0.37%	−1.13%***	−1.61%***	−2.04%***
	(0.83)	(−1.25)	(−0.88)	(−3.03)	(−4.14)	(−3.20)
5	2.52%***	1.79%***	1.90%***	2.84%***	3.52%***	1.00%*
	(6.21)	(4.33)	(5.00)	(7.08)	(9.62)	(1.82)
5 – 1	2.09%***	2.32%***	2.27%***	3.97%***	5.13%***	3.04%***
	(3.16)	(3.19)	(3.99)	(7.25)	(9.62)	(3.62)

Panel B: Portfolios formed by UE and $\hat{\rho}^2$

UE Quintile	$\hat{\rho}^2$ *Quintile*					
	1	2	3	4	5	5 − 1
1	0.78%	−0.59%	−0.78%*	−1.19%***	−1.44%***	−2.22%***
	(1.54)	(−1.30)	(−1.93)	(−3.15)	(−3.70)	(−3.52)
5	2.58%***	2.00%***	1.92%***	2.52%***	3.50%***	0.93%
	(5.97)	(4.82)	(5.00)	(6.88)	(9.51)	(1.64)
5 – 1	1.79%***	2.59%***	2.70%***	3.71%***	4.94%***	3.15%***
	(2.68)	(4.21)	(4.84)	(7.05)	(9.23)	(3.72)

Panel C: Portfolios formed by UE and $\hat{\rho}^2$						
			When $\hat{k}$ *is small*			
UE Quintile	$\hat{\rho}^2$ *Quintile*					
	1	2	3	4	5	5 − 1
5 – 1	2.70%***	1.33%	1.65%*	4.47%***	3.97%***	1.27%
	(3.21)	(1.60)	(1.74)	(4.24)	(3.09)	(0.65)
			When $\hat{k}$ *is large*			
UE Quintile	$\hat{\rho}^2$ *Quintile*					
	1	2	3	4	5	5 − 1
5 – 1	0.39%	3.47%***	2.88%***	4.58%***	5.28%***	4.89%***
	(0.26)	(3.01)	(3.28)	(5.98)	(8.17)	(3.08)

***Significant at the 1% level. *Significant at the 10% level.

Table 6
Predictive power of $\hat{\theta}_{iq}$ and $\hat{\rho}_{iq}^2$ relative to that of other firm/stock attributes

This table shows the results of the following regression model:

$$BHAR_{iq} = \beta_0 + \Psi_{iq} \times UE_{iq} + \sum_{n=1}^{10} \beta_{n+13} IndustryDummy_{n,iq} + \varepsilon_{iq},$$

$$\Psi_{iq} = \beta_1 + \beta_2 \hat{\theta}_{iq} + \beta_3 \hat{\rho}_{iq}^2 + \beta_4 \hat{k}_{iq} + \beta_5 \rho_{iq}^2 \hat{k}_{iq} + \beta_6 PI_{iq} + \beta_7 ESPREAD_{iq} + \beta_8 OD_{iq}$$
$$+ \beta_9 MVE_{iq} + \beta_{10} NAF_{iq} + \beta_{11} NII_{iq} + \beta_{12} TURN_{iq} + \beta_{13} PIN_{iq};$$

where the regression coefficient on *UE* is the estimate of *PEAD*, i.e., $PEAD_{iq} = \Psi_{iq}$. Panel A reports the results when we include *UE* and the interaction variable between *UE* and each firm/stock attribute in the regression. Panel B shows the results when we include these firm attributes, $\hat{\theta}_{iq} \times UE_{iq}$, and $\hat{\rho}_{iq}^2 \times UE_{iq}$ in the regression. The last two columns in Panel B show the results when we add two interaction variables (i.e., $\hat{k}_{iq} \times UE_{iq}$ and $\hat{\rho}_{iq}^2 \times \hat{k}_{iq} \times UE_{iq}$) in the regression. *UE* is earnings surprise, *BHAR* is the 60-day [+2, +61] size and B/M adjusted return, *PI* is the two-day [0, +1] adverse selection component/price impact measure, *ESPREAD* is the trade-weighted effective spread during the two-day event period, *OD* is the measure of opinion divergence, *MVE* is the market value of equity, *NAF* is the number of analysts following the firm, *NII* is the number of institutional investors that hold the firm's shares, *TURN* is the average daily turnover ratio during the 50-day period prior to earnings announcement, and *PIN* is the probability of informed trading. Appendix C provides details regarding the construction of these variables. Standard errors are adjusted for clustering by firm and time. Figures in parentheses are *t*-statistics.

Panel A: The relation between PEAD and firm/stock attributes

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
UE	0.02***	0.02***	0.03***	0.02***	0.02***	0.02***	0.03***	0.03***
	(8.24)	(8.34)	(8.41)	(7.24)	(7.05)	(7.23)	(8.54)	(9.40)
PI×UE	0.50							
	(0.97)							
ESPREAD×UE		2.28***						
		(4.29)						
OD×UE			−1.17**					
			(−2.36)					
MVE×UE				−3.07***				
				(−6.45)				
NAF×UE					−3.64***			
					(−7.63)			
NII×UE						−3.82***		
						(−8.15)		
TURN×UE							−2.86***	
							(−6.00)	
PIN×UE								2.76***
								(4.96)
Industry	No	No	No	No	No	No	No	No
R^2	0.002	0.002	0.002	0.002	0.002	0.003	0.002	0.003
N	106,253	106,443	106,440	106,443	106,443	106,201	106,440	70,707

Panel B: The relation between PEAD and $\hat{\theta}$ *and* $\hat{\rho}^2$								
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
UE	0.01^{***}	0.01^{**}	0.01	0.01^{***}	0.01^{**}	0.01	0.04^{***}	0.03^{**}
	(2.93)	(2.21)	(1.23)	(2.67)	(2.19)	(1.15)	(3.43)	(2.52)
$\hat{\theta}\times UE$	0.75^{***}	0.60^{***}	0.58^{***}					
	(5.92)	(3.83)	(3.38)					
$\hat{\rho}^2\times UE$				1.99^{***}	1.45^{***}	1.47^{***}	−1.11	−0.48
				(5.90)	(3.57)	(3.32)	(−1.28)	(−0.52)
$\hat{k}\times UE$							-0.16^{**}	-0.16^{*}
							(−2.22)	(−1.88)
$\hat{\rho}^2\times\hat{k}\times UE$							10.92^{***}	8.12^{**}
							(3.03)	(2.11)
PI×UE		−0.89	−0.53		−0.84	−0.51	−0.73	−0.41
		(−1.56)	(−0.86)		(−1.48)	(−0.83)	(−1.28)	(−0.66)
ESPREAD×UE		−1.07	-1.60^{*}		−1.03	−1.56	−0.71	−1.37
		(−1.25)	(−1.65)		(−1.21)	(−1.61)	(−0.85)	(−1.44)
OD×UE		−0.51	−0.47		−0.25	−0.25	1.15	1.27
		(−0.78)	(−0.59)		(−0.40)	(−0.33)	(1.08)	(1.00)
MVE×UE		0.53	−0.16		0.51	−0.21	0.43	−0.40
		(0.35)	(−0.10)		(0.33)	(−0.13)	(0.29)	(−0.24)
NAF×UE		-1.40^{*}	−1.41		-1.49^{*}	−1.46	-1.88^{**}	-1.96^{**}
		(−1.80)	(−1.51)		(−1.92)	(−1.57)	(−2.33)	(−2.07)
NII×UE		-3.03^{*}	−3.18		-3.08^{*}	−3.10	-3.16^{*}	−3.00
		(−1.77)	(−1.60)		(−1.80)	(−1.59)	(−1.90)	(−1.62)
TURN×UE		−0.79	−1.04		−0.84	−1.06	-1.23^{*}	-1.56^{*}
		(−1.24)	(−1.31)		(−1.31)	(−1.35)	(−1.81)	(−1.87)
PIN×UE			0.98			0.98		1.21^{*}
			(1.55)			(1.55)		(1.93)
Industry	No	Yes	Yes	No	Yes	Yes	Yes	Yes
R^2	0.002	0.056	0.059	0.002	0.056	0.059	0.056	0.059
N	104,996	104,996	69,511	104,996	104,996	69,511	104,996	69,511

***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

Table 7
Profitability of trading strategies based on $\hat{\theta}$, $\hat{\rho}^2$, and $\hat{k}$

We estimate $\hat{\rho}_{iq}$, $\hat{k}_{iq}$, and $\hat{\theta}_{iq}$ for all firms that announce earnings during January 2000 to December 2002 and calculate their mean values ($\hat{\rho}_i$, $\hat{k}_i$, and $\hat{\theta}_i$) for each firm during the period. Then, we divide all firms that announce earnings from October 2002 to December 2002 into five groups according to *UE* values (raw values before conversion to group numeric values). We also divide these firms into five groups according to $\hat{\theta}_i$, $\hat{\rho}_i^2$, and $\hat{\rho}_i^2 \times \hat{k}_i$, respectively. For each of the five groups formed by $\hat{\theta}_i$, $\hat{\rho}_i^2$, or $\hat{\rho}_i^2 \times \hat{k}_i$, we form a hedge portfolio that is long in the highest *UE* quintile firms and short in the lowest *UE* quintile firms and hold each of the five hedge portfolios for the month of January 2003. This completes the first rolling window. We repeat the above procedure and calculate the return of the five hedge portfolios for the month of February 2003. The last portfolio holding month is December 2012. Using the monthly returns of the five hedge portfolios, we calculate the Pastor-Stambaugh four-factor alpha for each hedge portfolio. This table reports the annualized alpha of each of the five hedge portfolios. Figures in parentheses are *t*-statistics.

Panel A: $\hat{\theta}$ Quintile					
1	2	3	4	5	5 − 1
0.52%	2.06%	4.73%	8.02%**	8.34%***	7.82%**
(0.15)	(0.63)	(1.57)	(2.57)	(3.19)	(2.23)

Panel B: $\hat{\rho}^2$ Quintile					
1	2	3	4	5	5 − 1
0.38%	3.89%	5.07%*	6.43%**	8.62%***	8.25%**
(0.11)	(1.17)	(1.71)	(1.96)	(3.26)	(2.33)

Panel C: $\hat{\rho}^2 \times \hat{k}$ Quintile					
1	2	3	4	5	5 − 1
−3.19%	3.65%	6.46%**	7.62%**	8.44%***	11.64%***
(−0.88)	(1.20)	(2.17)	(2.46)	(2.88)	(3.52)

***Significant at the 1% level. **Significant at the 5% level. *Significant at the 10% level.

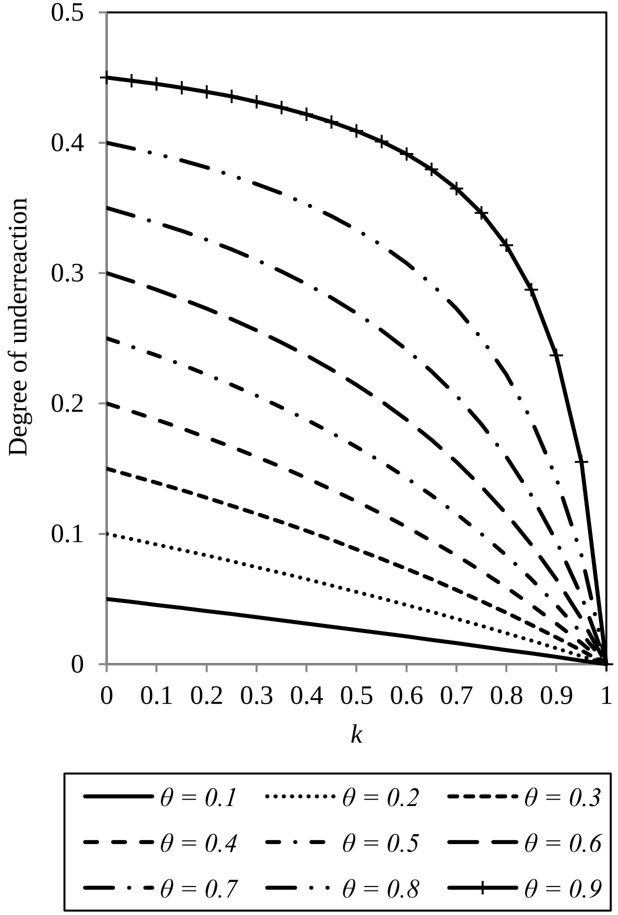

Figure 1a. Relation between the degree of underreaction and k at different levels of θ

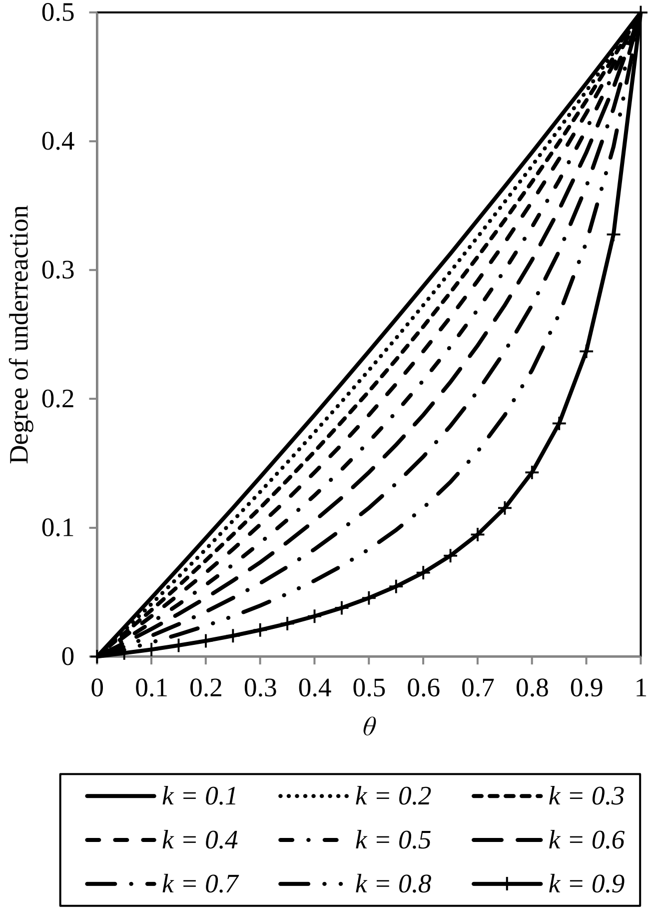

Figure 1b. Relation between the degree of underreaction and θ at different levels of k

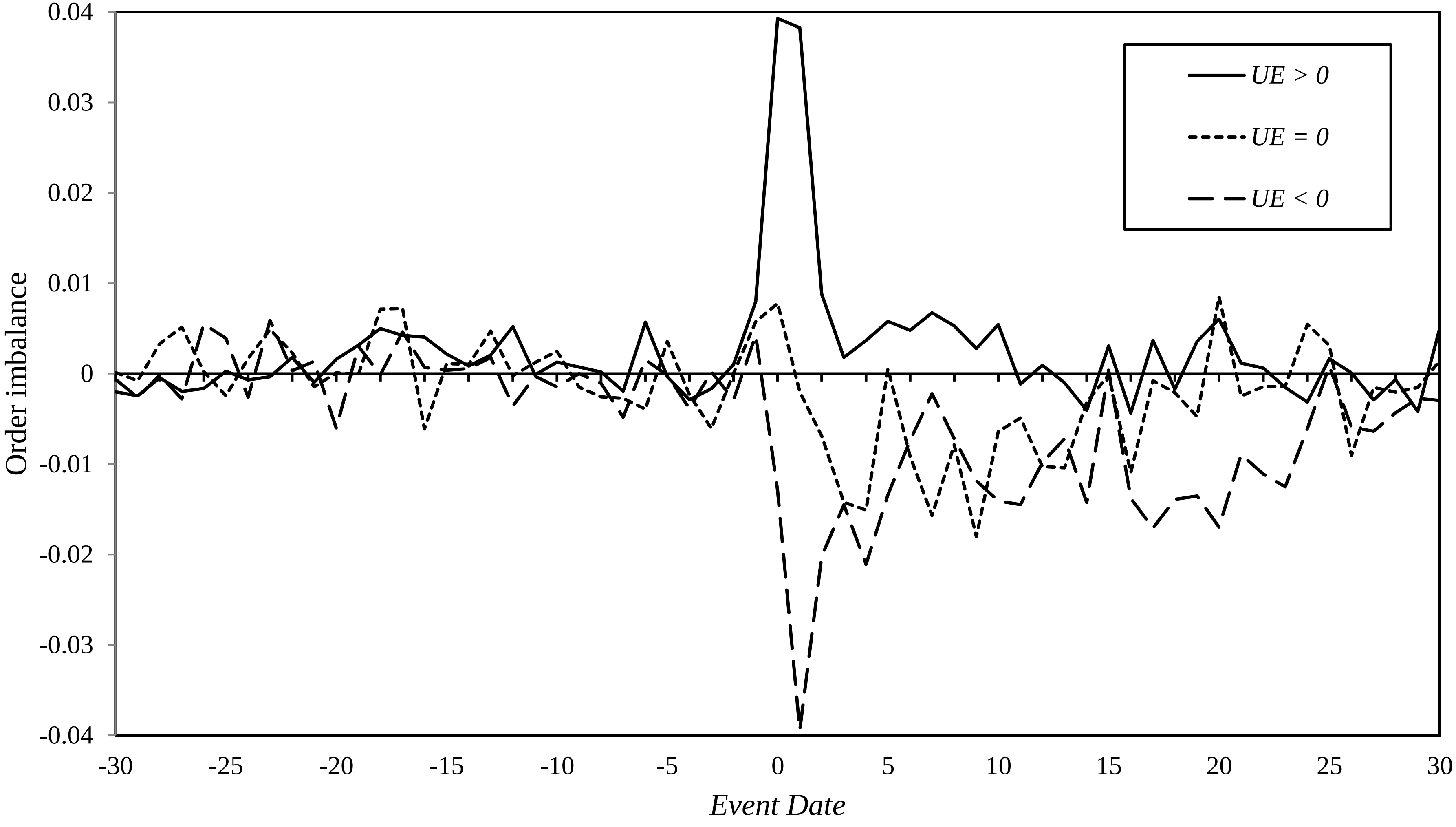

Figure 2. Daily order imbalance around earnings announcement

Appendix A
Derivation of *k*, *ERC*, and ρ equations
From:

$$\tilde{P}_2 = \left[\frac{m}{h+m} + \frac{s}{2(h+m+s)} \cdot \frac{h}{h+m}\right] \cdot \tilde{u} + \left[1 - \frac{s}{2(h+m+s)}\right] \cdot \frac{m}{h+m} \cdot \tilde{\eta} + \frac{s}{2(h+m+s)} \cdot \tilde{\varepsilon} + \frac{1}{2}\sqrt{\frac{s}{(h+m)(h+m+s)}}\tilde{\ell},$$

$$\tilde{\omega} = \sqrt{\frac{(h+m)sL}{h+m+s}} \cdot \left[\frac{h}{h+m}\tilde{u} - \frac{m}{h+m}\tilde{\eta} + \tilde{\varepsilon}\right] + \tilde{\ell},$$

and $\tilde{x} = \tilde{u} + \frac{m}{m+s}\tilde{\eta} + \frac{s}{m+s}\tilde{\varepsilon}$, it is straightforward to show that:

$$Cov[\tilde{\omega}, \tilde{x}] = \sqrt{\frac{(h+m)sL}{h+m+s}} \cdot \left[\frac{h}{h+m} \cdot \frac{1}{h} - \frac{m}{h+m} \cdot \frac{m}{m+s} \cdot \frac{1}{m} + \frac{s}{m+s} \cdot \frac{1}{s}\right]$$
$$= \sqrt{\frac{(h+m)sL}{h+m+s}} \cdot \left[\frac{h+m+s}{(h+m)(m+s)}\right] = \sqrt{\frac{(h+m+s)sL}{(h+m)(m+s)^2}},$$

$$Cov[\tilde{u}, \tilde{x}] = \frac{1}{h}$$

$$Var[\tilde{\omega}] = 2L$$

$$Var[\tilde{x}] = \frac{1}{h} + \frac{m}{(m+s)^2} + \frac{s}{(m+s)^2} = \frac{h+m+s}{h(m+s)},$$

and

$$Cov[\tilde{P}_2, \tilde{x}] = \left[\frac{m}{h+m} + \frac{s}{2(h+m+s)} \cdot \frac{h}{h+m}\right] \cdot \frac{1}{h} + \left[1 - \frac{s}{2(h+m+s)}\right] \cdot \frac{m}{h+m} \cdot \frac{m}{m+s} \cdot \frac{1}{m} + \frac{s}{2(h+m+s)} \cdot \frac{s}{m+s} \cdot \frac{1}{s}$$
$$= \frac{m}{(h+m)h} + \frac{s}{2(h+m+s)(h+m)} + \frac{m}{(h+m)(m+s)} - \frac{sm}{2(h+m+s)(h+m)(m+s)} + \frac{s}{2(h+m+s)(m+s)}$$
$$= \frac{m(m+s)+mh}{h(h+m)(m+s)} + \frac{s(m+s) - sm + s(h+m)}{2(h+m+s)(h+m)(m+s)} = \frac{2m(h+m+s)+hs}{2h(h+m)(m+s)}$$

From the above equations we get:

$$k \equiv \frac{Cov[\tilde{u},\tilde{x}]}{Var[\tilde{x}]} = \frac{1}{h} \cdot \frac{h(m+s)}{h+m+s} = \frac{m+s}{h+m+s}$$

$$\rho \equiv Corr[\tilde{\omega},\tilde{x}] = \frac{Cov[\tilde{\omega},\tilde{x}]}{\sqrt{Var[\tilde{\omega}]}\sqrt{Var[\tilde{x}]}}$$
$$= \sqrt{\frac{(h+m+s)sL}{(h+m)(m+s)^2}} \cdot \sqrt{\frac{1}{2L}} \cdot \sqrt{\frac{h(m+s)}{h+m+s}} = \sqrt{\frac{hs}{2(h+m)(m+s)}}$$

$$ERC = \frac{Cov[\tilde{P}_2,\tilde{x}]}{Var[\tilde{x}]} = \frac{2m(h+m+s)+hs}{2h(h+m)(m+s)} \cdot \frac{h(m+s)}{h+m+s} = \frac{2m(h+m+s)+hs}{2(h+m)(h+m+s)}$$
$$= \frac{1}{2}\left(\frac{m(h+m+s)}{(h+m)(h+m+s)} + \frac{m(h+m+s)+hs}{(h+m)(h+m+s)}\right) = \frac{1}{2}\left(\frac{m+s}{h+m+s} + \frac{m}{h+m}\right)$$

To express *ERC* and ρ as functions of k (information content) and θ (information asymmetry), we use the following identities:

$$k \equiv \frac{m+s}{h+m+s}, \ \theta \equiv \frac{s}{m+s}.$$

From the above two definitions we get:

$$k(1-\theta) = \frac{m}{h+m+s}, \ 1-k\theta = \frac{h+m}{h+m+s}, \text{ and thus, } \frac{k(1-\theta)}{1-k\theta} = \frac{m}{h+m}.$$

Now ρ and *ERC* can be rewritten as:

$$\rho = \sqrt{\frac{hs}{2(h+m)(m+s)}} = \sqrt{\frac{\theta}{2}\left(\frac{1-k}{1-k\theta}\right)},$$

$$ERC = \frac{1}{2}\left(\frac{m+s}{h+m+s} + \frac{m}{h+m}\right) = \frac{1}{2}\left(k + \frac{k(1-\theta)}{1-k\theta}\right) = k \cdot \frac{2-(1+k)\theta}{2-2k\theta}.$$

Appendix B
Derivation of *PEAD*

From

$$\tilde{u}-\tilde{P}_2=\left[1-\frac{m}{h+m}-\frac{s}{2(h+m+s)}\cdot\frac{h}{h+m}\right]\cdot\tilde{u}+\left[\frac{s}{s(h+m+s)}-1\right]\cdot\frac{m}{h+m}\cdot\tilde{\eta}$$
$$-\frac{s}{2(h+m+s)}\cdot\tilde{\varepsilon}-\frac{1}{2}\sqrt{\frac{s}{(h+m)(h+m+s)}}\tilde{\ell}$$

and $\tilde{x}=\tilde{u}+\frac{m}{m+s}\tilde{\eta}+\frac{s}{m+s}\tilde{\varepsilon}$, we have

$$Cov\left[\tilde{u}-\tilde{P}_2,\tilde{x}\right]=\left[1-\frac{m}{h+m}-\frac{s}{2(h+m+s)}\cdot\frac{h}{h+m}\right]\cdot\frac{1}{h}+\left[\frac{s}{s(h+m+s)}-1\right]\cdot\frac{m}{h+m}\cdot\frac{m}{m+s}\cdot\frac{1}{m}$$
$$-\frac{s}{2(h+m+s)}\cdot\frac{s}{m+s}\cdot\frac{1}{s}$$
$$=\frac{1}{h+m}\left[1-\frac{s}{2(h+m+s)}\right]-\frac{m}{(h+m)(m+s)}\left[1-\frac{s}{2(h+m+s)}\right]-\frac{s}{2(h+m+s)(m+s)}$$
$$=\left[1-\frac{s}{2(h+m+s)}\right]\cdot\left[\frac{m+s-m}{(h+m)(m+s)}\right]-\frac{s}{2(h+m+s)(m+s)}$$
$$=\frac{s}{(h+m)(m+s)}-\frac{s^2}{2(h+m+s)(h+m)(m+s)}-\frac{s}{2(h+m+s)(m+s)}$$
$$=\frac{2s(h+m+s)-s^2-s(h+m)}{2(h+m+s)(h+m)(m+s)}=\frac{2s(h+m+s)-s(h+m+s)}{2(h+m+s)(h+m)(m+s)}=\frac{s}{2(h+m)(m+s)}$$

$$PEAD\equiv\frac{Cov\left[\tilde{u}-\tilde{P}_2,\tilde{x}\right]}{Var\left[\tilde{x}\right]}=\frac{s}{2(h+m)(m+s)}\cdot\frac{h(m+s)}{h+m+s}=\frac{sh}{2(h+m)(h+m+s)}$$

Since $k\theta=\frac{s}{h+m+s}$ and $1-\frac{k(1-\theta)}{1-k\theta}=1-\frac{m}{h+m}=\frac{h}{h+m}$, it follows that

$$PEAD\equiv\frac{Cov\left[\tilde{u}-\tilde{P}_2,\tilde{x}\right]}{Var\left[\tilde{x}\right]}=\frac{1}{2}k\theta\cdot\left[1-\frac{k(1-\theta)}{1-k\theta}\right]=\frac{1}{2}k\theta\cdot\left[\frac{1-k\theta-k+k\theta}{1-k\theta}\right]=k\cdot\left[\frac{\theta}{2}\frac{1-k}{(1-k\theta)}\right].$$

Appendix C
Variable definitions and measurement

(a) *AR* is the two-day [0, +1] size and book-to-market ratio (B/M) adjusted return and is defined as follows:

$$AR_{iq} = \prod_{d=0}^{1}(1+ret_{iq}) - \prod_{d=0}^{1}(1+ret_{ff25})$$

where ret_{ff25} is the Fama-French 25 portfolio returns formed on size and the book-to-market ratio.

(b) *BHAR* is the 60-day [+2, +61] size and B/M adjusted return and is defined as follows:

$$BHAR_{iq} = \prod_{d=2}^{61}(1+ret_{iq}) - \prod_{d=2}^{61}(1+ret_{ff25})$$

where ret_{ff25} is the Fama-French 25 portfolio returns formed on size and the book-to-market ratio.

(c) *U* is the 62-day [0, +61] size and B/M adjusted return and is defined as follows:

$$U_{iq} = \prod_{d=0}^{61}(1+ret_{iq}) - \prod_{d=0}^{61}(1+ret_{ff25})$$

where ret_{ff25} is the Fama-French 25 portfolio returns formed on size and the book-to-market ratio.

(d) *PI* is the price impact a trade. $PI_{iq,d,t}$ of the trade at time *t* for firm-quarter *iq* on event date *d* is defined as follows:

$$PI_{iq,d,t} = 2D_{iq,d,t} \times \left(\frac{Mid_{iq,d,t+5} - Mid_{iq,d,t}}{Mid_{iq,d,t}} \right).$$

where $Mid_{iq,d,t}$ is the quote midpoint at time *t*, $Mid_{iq,d,t+5}$ is the first available quote midpoint between 5 minutes and 10 minutes after the trade time *t*, and $D_{iq,d,t}$ is equal to 1 for buyer-initiated trades and −1 for seller-initiated trades. We calculate the trade-weighted price impact over the two-day event window [0, +1].

(e) *ESPREAD* (the trade-weighted effective spread) is estimated over the two-day [0, +1] event window. The effective spread of trade at time *t* for firm-quarter *iq* on event date *d* is defined as follows:

$$ESPREAD_{iq,d,t} = 2D_{iq,d,t} \times \left(\frac{P_{iq,d,t} - Mid_{iq,d,t}}{Mid_{iq,d,t}} \right)$$

where $P_{iq,d,t}$ is the trade price at time *t*, $Mid_{iq,d,t}$ is the quote midpoint at time *t*, and $D_{iq,d,t}$ is equal to 1 for buyer-initiated trades and −1 for seller-initiated trades. The trade-weight average of the effective spread at each trade time is then calculated over the two-day [0, +1] event window.

(f) *PIN* is the probability of information-based trading measure [see Easley, Kiefer, O'Hara, and Paperman (EKOP), 1996]. We use Ellis, Michaely, and O'Hara (2000) algorithm to determine the number of buyer-initiated trade (B_d) and seller-initiated trade (S_d) for each day over the 40-day event window [−41, −2] prior to each earnings announcement. The daily likelihood function in EKOP is defined as follows.

$$L(\alpha,\delta,\mu,\varepsilon \mid (B_d,S_d)) = (1-\alpha)e^{-\varepsilon T}\frac{(\varepsilon T)^{B_d}}{B_d!}e^{-\varepsilon T}\frac{(\varepsilon T)^{S_d}}{S_d!}$$
$$+\alpha\delta e^{-\varepsilon T}\frac{(\varepsilon T)^{B_d}}{B_d!}e^{-(\mu+\varepsilon)T}\frac{[(\mu+\varepsilon)T]^{S_d}}{S_d!},$$
$$+\alpha(1-\delta)e^{-(\mu+\varepsilon)T}\frac{[(\mu+\varepsilon)T]^{B_d}}{B_d!}e^{-\varepsilon T}\frac{(\varepsilon T)^{S_d}}{S_d!}$$

where α is the probability that an information event has occurred, δ is the probability of a low signal given that an event has occurred, μ is the probability that a trade comes from an informed trader given that an event has occurred, ε is the probability that uninformed traders will actually trade, T is the total trading time for the day. We estimate $\alpha,\delta,\mu,\varepsilon$ by maximizing the product of daily likelihood function,

$$L(\alpha,\delta,\mu,\varepsilon \mid M) = \prod_{d=-41}^{-2} L(\alpha,\delta,\mu,\varepsilon \mid B_d,S_d).$$

PIN measure is then defined as $PIN = \dfrac{\hat{\alpha}\hat{\mu}}{\hat{\alpha}\hat{\mu}+2\hat{\varepsilon}}$.

(g) *OD* is a proxy for opinion divergence at the time of earnings announcement (see Garfinkel et al., 2006). OD_{iq} for firm-quarter *iq* is defined as follows:

$$OD_{iq} = \frac{\left\{\sum_{d=0}^{1}\left[\left(\frac{Vol}{Shrout}\right)_{iq,d} - \left(\frac{Vol}{Shrout}\right)_{mkt,d}\right]\right\}}{2} - \frac{\left\{\sum_{d=-54}^{-5}\left[\left(\frac{Vol}{Shrout}\right)_{iq,d} - \left(\frac{Vol}{Shrout}\right)_{mkt,d}\right]\right\}}{50}$$

where $d = 0$ is the earnings announcement date for calendar quarter q and $\left(\frac{Vol}{Shrout}\right)_{iq,d}$ is share turnover ratio on event date d.

(h) *MVE* (market value of equity) is the product of closing price and number of shares outstanding at fiscal quarter end.

(i) *NAF* is the number of analysts following the firm during the one-year period prior to earnings announcement.

(j) *NII* is the number of institutional investors holding the share of the firm prior to the fiscal quarter end.

(k) *TURN* (turnover ratio) is the daily ratio of share volume to number of shares outstanding averaged across the 50-day period [−54, −5] prior to earnings announcement.

(l) *BETA* (systematic risk) is the systematic risk of a stock, which is the slope coefficient obtained from the regression of daily excess return on daily excess market return over the 250-day event window prior to the earnings announcement.

(m) *SIGMA* (unsystematic risk) is the standard deviation of the residuals in the regression model for BETA.

(n) *EAPER* (earnings persistence) is estimate of φ in the following regression:

$$E_{iq}/A_{iq} - E_{iq-4}/A_{iq-4} = \varphi\left(E_{iq-1}/A_{iq-5} - E_{iq-5}/A_{iq-5}\right) + \varepsilon_{iq}$$

where E_{iq}/A_{iq} is net income before extraordinary items divided by total asset for firm *i* in calendar quarter *q*.

(o) *MBR* (market-to-book ratio) is the ratio of the market value of the firm (the book value of total assets – the book value of equity + market value of equity) to the book value of total assets at the end of the previous fiscal year.

(p) *LOSS* is equal to 1 if earnings at the current fiscal quarter end are negative and 0 otherwise.

(q) *EAVOL* (earnings volatility) is the standard deviation of the difference in earnings (net income before the extraordinary items divided by total assets) between the most recent fiscal quarter end and the fiscal quarter end four quarters ago using the eight-year period of data prior to earnings announcement.

(r) *PREANN* is the 40-day [−41, −2] preannouncement return and is defined as follows:

$$PREANN_{iq} = \prod_{d=-41}^{-2}(1+ret_{iq}) - \prod_{d=-41}^{-2}(1+ret_{ff25})$$

where ret_{ff25} is the Fama-French 25 portfolio returns formed on size and book-to-market ratio.