Three essays on information risk

Qian, Xiao Lin

2010

Qian, X. L. (2010). Three essays on information risk. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/20859>

<https://doi.org/10.32657/10356/20859>

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

THREE ESSAYS ON INFORMATION RISK

**THREE ESSAYS ON INFORMATION
RISK**

XIAOLIN QIAN

XIAOLIN QIAN

NANYANG BUSINESS SCHOOL

2010

2010

Three Essays on Information Risk

Xiaolin Qian

Nanyang Business School

A thesis submitted to the Nanyang Technological University
in partial fulfilment of the requirement for the degree of

Doctor of Philosophy

2010

ACKNOWLEDGEMENTS

I would like to express my sincere gratitude to Professor Chuan-Yang Hwang, for supervising me in my PhD study and research. His patience, motivation, enthusiasm, and immense knowledge helped me in all the time of research and writing of this thesis.

Special thanks are also given to Professor Sie Ting Lau and Professor Lilian Ng, for serving as my dissertation committee. I am very fortunate to have received their kind assistance and valuable advice.

I would also like to thank all the professors in the Division of Banking and Finance who were ever willing to teach or encourage me. Furthermore, I would like to acknowledge the generous help from Ms. Amarnisha Mohamed throughout my doctoral study.

I dedicate this dissertation to my parents and my husband for their encouragement and endless love.

TABLE OF CONTENTS

Acknowledgements	1
List of Tables	4
List of Figures.....	5
SUMMARY	6
1. IS INFORMATION RISK PRICED? EVIDENCE FROM THE PRICE DISCOVERY OF LARGE TRADES	9
1.1 Introduction.....	9
1.2 Data and Sample	22
1.2.1 Large Trades and Small Trades Classification.....	24
1.2.2 Price Impact Analysis of Large and Small Trades.....	29
1.3 VECM and the Construction of ECIN	31
1.4 Empirical Results	38
1.4.1 The Relationship of ECIN with Traditional Information Asymmetry Measures.....	41
1.4.2 Asset Pricing Test	42
1.5 Conclusion	56
2. INFORMATION RISK FACTOR AND ASSET PRICING ANOMALIES.....	72
2.1 Introduction.....	72
2.2 Data and Estimation of ECIN	82
2.3 ECIN Portfolios and ECIN Factor	84
2.4 Asset Pricing Anomalies	91
2.4.1 Price Momentum.....	92
2.4.2 Post-Earnings-Announcement Drift.....	94
2.4.3 Long-Run Underperformance after SEO and Stock Issuance.....	97
2.5 Conclusion	103
3. THE EFFECT OF REGULATION FAIR DISCLOSURE (REG FD) ON INFORMATION ASYMMETRY AND THE COST OF CAPITAL: REVISITED.....	121
3.1 Introduction.....	121
3.2 Proxy for Information Asymmetry and Proxies for Information Dissemination	127
3.3 Data and Sample	128
3.4 Asset Pricing Tests.....	131

3.5 Changes in Information Asymmetry and in the Cost of Capital around Regulation Fair Disclosure	133
3.6 Conclusion	140
References.....	152

LIST OF TABLES

Table 1-1 Summary Statistics	58
Table 1-2 Large and Small Trades	59
Table 1-3 Temporary and Permanent Price Effects of Large and Small Trades.....	60
Table 1-4 VECM Parameters	62
Table 1-5 Characteristics of <i>ECIN</i>	64
Table 1-6 Asset Pricing Tests	66
Table 1-7 Asset Pricing Tests—Further Checks	67
Table 1-8 Asset Pricing Tests – Comparing <i>ECIN</i> with <i>PIN</i>	68
Table 1-9 Economic Significance of <i>ECIN</i>	69
Table 2-1 Characteristics and Returns of <i>ECIN</i> Portfolios	106
Table 2-2 Return Correlations between <i>ECIN</i> -Based Zero-Investment Portfolios.....	108
Table 2-3 Descriptive Statistics and Time-Series Relations of Risk Factors.....	109
Table 2-4 Portfolio Time-Series Regressions of Contemporaneous Excess Returns on Factor Returns	110
Table 2-5 Monthly Cross-Sectional Regressions of Returns on Factor Loadings	112
Table 2-6 Momentum Portfolios on NYSE/Amex	114
Table 2-7 Momentum Portfolios on NYSE/Amex/Nasdaq	115
Table 2-8 Post-Earnings-Announcement Drift on NYSE/Amex	116
Table 2-9 Post-Earnings-Announcement Drift on NYSE/Amex/Nasdaq	117
Table 2-10 Seasoned Equity Offerings Portfolios	118
Table 2-11 Time Series Regressions of Share Issue Portfolios	119
Table 3-1 Summary Statistics	142
Table 3-2 Correlations between <i>ECIN</i> and Information Dissemination Proxies	143
Table 3-3 Fama and MacBeth Regression Results	144
Table 3-4 Fama and MacBeth Regression Results with the Illiquidity Measure.....	145
Table 3-5 Fama and MacBeth Regression Results with <i>ECIN</i> -Squared.....	146
Table 3-6 Changes in <i>ECIN</i> and Information Dissemination	147
Table 3-7 Firm Level Cross-Sectional Cost of Capital Changes	148
Table 3-8 Cost of Capital Changes and Firm Characteristics	151

LIST OF FIGURES

Figure 1-1 Cumulative Average Returns in Seller (Buyer)-Initiated Trades	71
Figure 2-1 Time Series of SEO Listed by Exchange at the Time of Offering.....	120

SUMMARY

This dissertation contains three essays. The first essay develops an information risk measure that is based on the price discovery of large trades. This is motivated by the observation that informed traders prefer to trade in large size, so the price discovery of large trades would yield a more precise measure of the information risk than would price discovery of small trades. As the price series of large trades and small trades are cointegrated, the price discovery of trades can be estimated via the vector error-correction model (VECM). Intuitively, we use VECM to study how a temporary gap between the large trade price and the small trade price of the same stock is going to be closed. If closing the gap is mostly done through the small trade price with little movement in the large trade price, the large trade price has been closer to the long-run equilibrium price, and the large trade price has a greater price discovery function for this stock. Using this new measure, we show that information risk is priced. The pricing impact of information risk is not a manifestation of liquidity effect as some have argued for the case of *PIN*. Moreover, it subsumes the pricing impact of both *PIN* and Amihud's (2002) *ILLIQ*. These results demonstrate the effectiveness of our information risk measure, and suggest that the liquidity effect captured by *ILLIQ* has its origin in the information asymmetry.

Essay two constructs an information risk factor as the return of the mimicking portfolio that takes the long position in the size-neutral high-*ECIN* portfolio and the short position in the size-neutral low-*ECIN* portfolio. The two-stage cross-sectional regression tests confirm that this is a priced risk factor. We further use it to investigate the following asset pricing anomalies: price momentum, earnings momentum (post-earnings-announcement drift), the long-term underperformance of

SEO firms, and the share issuance puzzle. We show that stocks with smaller or negative earnings surprise (losers in the price momentum) have significantly lower loading on the information risk factor than stocks with large earning surprise (winners in the price momentum.) The differences in the loadings on information risk can explain the price momentum and earning drift anomalies. We also show that both SEO firms and firms with positive share issuance experience a significant reduction in the information risk after SEO and share issuance. The lower and negative loadings on the information risk factor of the SEO and share issuance firms can resolve the long-run underperformance anomaly associated with the SEO and share issuance firms. This paper provides concrete evidence showing that the information risk factor is a significant price factor. Asset pricing tests that fail to incorporate it will potentially generate misleading anomalies.

Essay three examines the change in *ECIN* and the related cost of capital after Regulation Fair Disclosure (Reg FD). Using *PIN* as the information asymmetry (risk) measure, Duarte, Han, Harford and Young (2008) find that Reg FD increases information asymmetry and the cost of capital of Nasdaq firms but has no impact on NYSE/Amex firms. Their results suggest that Reg FD has overall negative effects on information environment. As Duarte and Young (2009) shows that *PIN* captures liquidity rather than information risk, we revisit the study by using *ECIN*, developed in the first essay that is an effective information risk measure whose price impact subsumes *PIN* and liquidity. Like Duarte, Han, Harford and Young (2008), we find that Reg FD increases information asymmetry and the cost of capital of Nasdaq firms. However, in sharp contrast with theirs study, we find that Reg FD is beneficial to NYSE/Amex firms that it reduces information asymmetry and lowers the cost of

capital. A further test reveals that the different impact of Reg FD on NYSE/Amex firms and on Nasdaq firms are because of the different firm characteristics. In particular, the change in the cost of capital is negatively related to firm size, book-to-market and *ECIN*.

1. IS INFORMATION RISK PRICED? EVIDENCE FROM THE PRICE DISCOVERY OF LARGE TRADES

1.1 Introduction

The effect of information asymmetry on stock returns has become the subject of intense debate in asset pricing literature in recent years. Information asymmetry plays no role in traditional asset pricing models like CAPM, APT and the consumption CAPM because symmetric information is assumed in these models. The argument that information asymmetry would be totally eliminated in a fully-revealing equilibrium, is often used to support the view that information asymmetry should not affect asset returns. However, Grossman and Stiglitz (1980) show that a fully-revealing equilibrium is only possible in the unlikely case where information is costless. Admati (1985) extends Grossman and Stiglitz's partially-revealing equilibrium model to speculative markets in which there are multiple risky-assets and traders with diverse and asymmetric information. In this extended model, each agent has a different risk-return tradeoff, depending on her private signal and the distribution of signal. Wang (1993) presents a model of the dynamic asset pricing under asymmetric information. In his model, information induces an adverse selection problem and uninformed traders demand a premium for the risk of trading with informed traders. However, trading by informed investors also makes prices more informative, thereby reducing uncertainty and hence reducing the return premiums. The overall effect of the asymmetric information on the required return is ambiguous. Garleanu and Pedersen

(2004) present a model that shows that adverse selection increases the required return through allocation inefficiency. Easley and O'Hara (2004) show in a multiple asset partially-revealing rational expectation model that due to information asymmetry, uninformed investors face information risk as they are not able to adjust their portfolio weights to take advantage of new information as well as informed investors do. Furthermore, this information risk is not diversifiable; therefore, uninformed investors require higher return as compensation to hold shares in firms with higher information asymmetry. On the other hand, Hughes, Liu and Liu (2007) and Lambert, Leuz and Verrecchia (2007) argue that the information risk is either fully diversifiable when economy is large enough or has been captured by existing risk measures. Most recently, Lambert et al (2008) show that in rational expectation models of perfect competition, information difference across investors affects a firm's cost of capital through investors' average information precision and not information asymmetry per se.

On the empirical front, Easley, Hvidkjaer and O'Hara (2002) use direct information risk measure and find evidence that information asymmetry affects asset returns. Relying on the structural microstructure model developed by Easley, Kiefer, O'Hara and Paperman (1996), they estimate the probability of information-based trade (*PIN*) directly from the trading data and find that stocks with higher *PIN* have higher expected returns. However, Duarte and Young (2009) find that *PIN* is priced because it captures the effect of illiquidity that is unrelated to the asymmetric information. When *PIN* is decomposed into two components, one related to asymmetric information and the other related to illiquidity, only the component related to illiquidity is priced. Given this evidence, one may be tempted to conclude that information risk is not priced. However, this is a premature conclusion, since there is

evidence of the limitation of *PIN* as the proxy of information risk. Aktas, de Bodt, Declerck and Van Oppens (2007) investigate *PIN* around a sample of merger and acquisition announcements in Euronext Paris and find that the behavior of *PIN* is inconsistent with the evidence of information leakages during the pre-event period. They argue that *PIN* fails to capture the information risk in their sample, and attribute the failure to the convergence problem encountered in the *PIN* estimation of the highly actively traded stocks. Aslan, Easley, Hvidkjaer and O'Hara (2007) are also aware of the convergence problem in their *PIN* estimation, stating that “in stocks with very large number of trades, the optimization program encounters computational underflow, and we are unable to evaluate the likelihood function for those stocks.” The sample of Aslan, Easley, Hvidkjaer and O'Hara (2007) consists of the US stocks between 1983 and 1999. We expect the convergence problem to be aggravated in more recent samples since the trading frequency and the trading volume have increased dramatically, especially after the decimalization of tick size in 2001. In addition, Easley, Hvidkjaer and O'Hara (2005) shows that while only 3.6% of the stocks in the year 2001 sample do not obtain *PIN* estimates, these stocks account for 23.7% of the total market capitalization. This result suggests *PIN* estimation works especially poorly in large stocks. This size related systematic estimation error of *PIN* can potentially induce bias in asset pricing tests.

In this paper we develop an alternative measure of information risk that avoids the convergence problem encountered in the *PIN* estimation. Our measure is estimated directly from the price discovery of large trades, and is motivated by the following observations. First, the private information is revealed in a sequence of

trade prices. Grossman (1976) models how an investor's private information about a risky asset's future payoff affects both her demand for that asset and the asset's equilibrium price. New information is incorporated into the market price and a rational individual can learn about others' private information from the risky asset's price. This process is known as "price discovery" described by Schreiber and Schwartz (1986). Our information risk measure is developed from the concept of the price discovery. The basic assumption is that only private information is revealed from the price discovery process, because public information is incorporated into the stock prices immediately without any impact on trades. Some may argue that if investors interpret public information differently, the intrinsic value also needs to be realized through trades or equivalently from the price discovery process. But we may notice that the public information in this scenario also introduces the information asymmetry, where the sophisticated investors trade in the direction towards the intrinsic value and the naïve investors trade in the opposite direction. Thus the public information has the "private components", which is the different interpretation of the public information. Because this information asymmetry also makes naïve investors lose to sophisticated investors, it is still under the study of information risk. Glosten and Milgrom (1985) and Kyle (1985) show that trades play a role in the price discovery as private information will be revealed gradually by informed trades. Subsequently, the price discovery of informed traders' trades reflects their information advantage. In particular, the more private information informed traders have, the larger price discovery their trades reveal. On the other hand, a small price discovery suggests that the magnitude of private information is small. And more importantly, the price discovery indicates that informed traders have utilized their superior information in trades, where the informed gains and the uninformed lose. According

to Easley and O'Hara (2004), uninformed traders accordingly require higher return to compensate such losing opportunity caused by the information asymmetry.

Second, informed and uninformed trades can be distinguished by the trade size. Since the price discovery of informed trades indicates the magnitude of information asymmetry risk, we first need to identify the informed trades. This is not easy because informed traders typically camouflage their trades in the uninformed trades. But researchers uncover some characteristics of the informed trading behavior. First, informed traders prefer to trade in large size if their private information is short-lived or if they trade competitively. Since informed traders aim to maximize the profit from private information, small trades are suboptimal given the relatively large proportion of fixed cost in each trade. The positive correlation between information and trade size is supported by the theoretical works of Easley and O'Hara (1987), Grundy and McNichols (1989), Holthausen and Verrecchia (1990), and Kim and Verrecchia (1991) and the empirical works of Madhavan and Smidt (1991), and Easley, Kiefer and O'Hara (1997). On the other hand, Kyle (1985) and Admati and Pfleiderer (1988) argue that the monopolist informed trader camouflage his trades by splitting one large trade into several small trades. The empirical work by Barclay and Warner (1993) confirms the prediction of these models that the informed traders do split their orders. However, Barclay and Warner (1993) also suggest that if informed traders break up trades to camouflage their private information, they may optimally submit medium-sized orders. For example, they find informed trades are concentrated in the medium-sized trading category from 1,000 to 9,900 shares. Such medium-sized trades account for 45.7% of all trades, but they reveal 92.8% of the total cumulative price changes.

Hasbrouck (1991a), Hausman, Lo, and MacKinlay (1992), and Kempf and Korn (1999) also find the positive but concave relation between price impact and trade size. More recently, Chan and Fong (2000) illustrate that NYSE stocks have the strongest volume-volatility relation in the medium size trades (5,000 to 9,999 shares.) That the block trades (10,000 shares and above) are more likely to be uninformed may attribute to the different market-making mechanism as discussed by Seppi (1990) and Hasbrouck (1991a). For example, most small or medium trades for NYSE firms will be executed in the specialist market anonymously, but most block trades are negotiated in the upstairs market where traders' identities are known. Therefore the medium-sized trades give informed investors better camouflage than the block trades. Consequently, informed traders prefer to submit orders of this medium size.

In this paper, we group trades into two categories, namely the potentially-informed-trades and the least-likely-informed trades. Based on the previously discussed literature, we classify the trades according to the trade size. Specially, the medium to large trades are in the potentially-informed-trades group, while the small trades are in the least-likely-informed-trades. Given the compelling evidence in literature that block trades (10,000 and above) are not likely to be informed, we test the price discovery of the large trades with or without block trades separately and compare the results. We find that the information risk measures do not have large difference under the two methods. In addition, the information risk measure is significantly priced in either case. This small influence of block trades might be due to the small percentage of block trades (less than 2% on average) in the total number of trades. But we also find the information risk measure does perform a little stronger by excluding block trades from the potentially-informed-trades group, consistent with the previous literatures that block trades tend to be uninformed. To improve the power

of identifying potentially-informed-trades, we exclude block trades from the potentially-informed-trades group. Since we only have two groups of trades classified by the trade size, we denoted them as large and small trades for the notation convenience.

We employ the dollar-based classification method, which conditional on both the stock price and the firm size. In addition, because of significant innovations in trading technology over the sample period, the trade-size criteria may be sensitive to the time period. We therefore use two subsamples to construct the trade-size criteria to capture the trading technology or market condition changes over time. But the results are actually robust if we use one or three sample periods. By dividing the trades into only two trade size groups (large and small), our classification of the large size trades covers most of the trades category in which previous papers have found the informed traders concentrate their trading. Moreover, with a similar classification method as we employed, Lee and Radhakrishna (2000) document that individual and institutional investors can be distinguished by their trade size, where the institutional investors are more likely to trade in large size. Battalio and Mendenhall (2005) and Hvidkjaer (2006) show that investors initiating large trades are more sophisticated while investors initiating small trades are more likely to respond to the naïve signals.

Although the trade size is an indicator of trading informativeness, we must note that not all medium or large trades are information-based trades. Investors can trade in medium or large size only for liquidity reasons. So we call these trades as the *potentially-informed-trades*. The price discovery of these trades assesses how much information advantage this group of trades has. If most medium to large trades are just

liquidity-motivated trades, we would observe little price discovery in this group. But if many trades are information-motivated, we would observe large price discovery in this group. Consequently, the price discovery of these trades estimates the information advantage or the information asymmetry between the potentially-informed trades and the uninformed small trades.

Third, the prices of large trades and small trades are cointegrated and can be described by the vector error-correction model (VECM) in which the error-correcting coefficients measure the degree of price discovery. The VECM allows one to examine the short-run dynamics in which each of the cointegrated series move from disequilibrium toward a long-run equilibrium. In our context, a long run equilibrium is one in which the large trade price equals the small trade price as they are the prices of the same stock. Intuitively, we use VECM to study how a temporary gap between the large trade price and small trade price of the same stock (a disequilibrium caused by information or liquidity reason) is going to be closed. If closing the gap is mostly done through the small trade price with little movement in the large trade price, this would result in a small (in absolute value) error-correcting coefficient in VECM for large trade price. This also suggests that the large trade price has been closer to the long-run equilibrium price, hence that the large trade price has a greater price discovery function for this particular stock. On the other hand, if the gap is closed mostly through the reversal movement of large trade price, then the gap in the previous period was more likely to be caused by liquidity reason and, in this sense, the large trade price has a small price discovery and a correspondingly large (in absolute value) error-correcting coefficient. Since large trades are potentially privately informed trades as argued above, the error-correction coefficient of large trades measures the price discovery hence the possibility and magnitude of private information-based trade, i.e.,

the information risk. As our information risk measure is based on the error-correction coefficient of potentially informed trades, it will be referred to as *ECIN* in the remaining part of the paper.

Armed with *ECIN*, we contribute to the information risk pricing debate by showing that information risk is indeed priced—stocks with higher *ECIN* earn higher returns. Moreover, this result is robust to beta, firm size, book-to-market ratio, illiquidity, and momentum effects. The pricing impact of *ECIN* cannot be explained by the standard risk factors in the literature. A strategy that takes a long position in the stocks in top *ECIN* quintile and takes a short position in the stocks in bottom *ECIN* quintile generates a significant four-factor (*MKT*, *HML*, *SMB* and *UMD*) risk-adjusted return of 0.29% to 0.39% per month.

ECIN dominates *PIN* in predicting returns. Although both *PIN* and *ECIN* have significantly positive coefficients when they are included separately in the return regressions, *PIN* ceases to be significant while *ECIN* remains significantly positive when they are jointly included. Furthermore, the pricing impact of *ECIN* cannot be explained by illiquidity. Duarte and Young (2009) illustrate that *PIN* is priced because it captures the effect of stock illiquidity. When the Amihud (2002) illiquidity measure (*ILLIQ*) is included in the asset pricing test, the significance of *PIN* disappears. In contrast, the pricing impact of *ECIN* not only survives the control of various illiquidity-related measures such as bid-ask spread, return volatility, stock turnover, it also subsumes the pricing impact of *ILLIQ*.

These results have a few important implications. First, they suggest that *ECIN* is a more effective information risk measure than *PIN*. Second, the liquidity effect that *ILLIQ* captures originates from information asymmetry. This is consistent with theoretical models of Kyle (1985), Glosten (1989) and Easley and O'Hara (1987). Brennan and Subrahmanyam (1996) also argue that liquidity effects of asymmetric information are most likely to be captured in the price impact of a trade, or the variable component of trading cost that *ILLIQ* is designed to capture. Third, these results suggest that information risk is not only priced but is priced beyond the liquidity effect of asymmetric information, and are consistent with theoretical models of O'Hara (2003) and Easley and O'Hara (2004). O'Hara (2003) argues that information asymmetry can affect required return via channels of liquidity and price discovery risk. The former channel is established through the fact that illiquidity affects traders' transaction costs and it remains active even when traders are risk-neutral. In contrast, the latter channel is established through the fact that risk of price discovery affects traders' risks hence their utilities, and it is active only when traders are risk averse. Complementary to our study, Tong (2009) uses abnormal large volumes in large trade to measure the intensity of information based trading. Consistent with our findings, he finds that an abnormal volume in large trade is positively priced in the cross-section after controlling for *ILLIQ*.

However, people may still argue that *ECIN* only capture certain form of illiquidity risk. The discussion of this issue actually depends on how illiquidity or liquidity is defined. Liquidity can be defined as how easily investors can trade without moving price, or it can also be defined as how easily investor can trade with the (new) efficient price. Under the perfect-revealing information assumption, these two definitions are identical. All information is public information and is incorporated into

stock price immediately. But if the financial market is not under informational perfection (and it is most likely not), price impact of a trade may come from either the trading friction or the new (private) information. If we define the liquidity as how easily investors can trade without moving prices, it is difficult or even impossible to distinguish these two effects. Therefore the illiquidity also captures the price impact from new information, which is highly related to the information risk. The illiquidity measures based bid-ask spread also contain a similar “information components.” Although theoretically we can distinguish the two functions of financial markets, i.e. providing liquidity and price discovery (O’Hara (2003)), they are not easily segmented empirically. When market makers quote high spread to compensate the loss from trading with informed traders, this cost is eventually imposed to the uninformed traders and increases the information risk. Therefore, the illiquidity measures simultaneously capture the information risk. But the illiquidity measures cannot distinguish market friction and information asymmetry. For example, an order can move the price a lot because it reveals superior information or just because the stock is illiquid. Market makers may quote large bid-ask spread because there is large probability of informed trading or because the illiquid stock have large inventory and order processing costs. But the information risk measure should capture only the information asymmetry risk, not the market friction. *ECIN* is based on the price discovery of large trades, which is the reaction of large trades to the disequilibrium (difference) between large and small trade prices. This disequilibrium can be caused by trading frictions or private information. If it is caused by trading frictions, we will observe a following reversal in the price of the trade that brings the disequilibrium in the first place. For example, at time t , both large and small trades are traded at the

price \$20, so the difference is zero. At time $t+1$, there is a large (small) buy, price is moved up to $\$20\frac{1}{8}$ for large (small) price only and the price of small (large) does not change. If the trade at time $t+1$ does not contain any information, the price movement is purely trading cost and we observe the large (small) price goes back to \$20 at time $t+2$ and still the small (large) price does not change. In this process, there is a reversal in the price which initiate the trade at time $t+1$. On the other hand, if the trade at $t+1$ reflect the new intrinsic value, the new equilibrium price at $t+2$ will reflect this information and we will observe both prices increase to $\$20\frac{1}{8}$. There is continuity in the price which initiated the trade at time $t+1$. The price discovery coefficients of VECM potentially capture the continuity or reversal in the large and small trade prices, since the less responds (movement) to the disequilibrium results in a higher price discovery. As we assume most informed traders optimally submit the medium- to large-sized orders, the price discovery of large trades reflects how the private information is revealed through trades. Information-based trades results in a small movement and a large price discovery coefficient for large trades. But illiquidity, on the other hand, leads to a large reversal following the uninformed large trades and a small price discovery coefficient for large trades. In addition, the informed large trade certainly incurs transaction costs as well. In most case of informed trades, price impact is a mix of both information and illiquidity. The price discovery can distinguish these two effects: the larger the price discovery (more continuity), the more information is revealed by the large trades; the smaller the price discovery (more reversal), the higher transaction costs and more trading frictions are faced by the informed large trades. Thus the price discovery of large trades actually measures how easily the private information can be traded and incorporated into stock prices, which should be different with the illiquidity risk. Because the private information in the

small trade is negligible, the price discovery coefficient of small trades may capture the price reversal from previous small trades or the reaction to the informed large trades. These two effects cannot be segmented easily to reflect the impact of private information, so we do not include the price discovery of small trades in the construction of information risk measure.

The contribution of this paper goes beyond developing a good information risk measure and showing that information risk is priced in the information risk pricing debate. *ECIN* will find great applications in corporate finance and accounting research. For example, our measure can be used to study the impact on the cost of capital of events that are hypothesized to affect the information production, such as the implementation of Regulation Fair Disclosure (Duarte, Han, Harford and Young (2008)). This more powerful information risk measure can also be used to study or re-examine the interactions between information asymmetry and 1) corporate governance (Ferreira and Laux (2007)), 2) regular conference (Brown, Hillegeist and Lo (2004)), 3) earning reporting outcome (Jayaraman (2008)), 4) conservatism in financial statements (LaFond and Watts (2008)), 5) post-earning announcement drift (Vega (2006)), and 6) firm investment and its sensitivity to cash flows (Chen, Goldstein and Jiang (2007) and Ascioglu, Hegde and McDermott (2008)).

Our results are also related to the growing literature of disclosure and the cost of capital. A large amount of finance and accounting literature (e.g. Botosan (1997), Botosan and Plumlee (2002), Botosan, Plumlee and Xie (2004), Francis, LaFond, Olsson and Schipper (2004), Aboody, Hughes and Liu (2005), Lambert, Leuz and Verrecchia (2006), and Francis, Nanda and Olsson (2008)) shows that a firm's quality

of information affects its cost of capital. Liquidity and estimation risk are commonly thought as two channels via which disclosure affects the cost of capital. Diamond and Verrecchia (1991) maintain that greater disclosure reduces the amount of information revealed by a large trade. When the adverse price impact of such trade is reduced, the corresponding increase in liquidity spurts the demand for a firm's securities from the large investors which in turn reduces its cost of capital. Barry and Brown (1985), Handa and Linn (1993), Coles, Loewenstein and Suay (1995), and Clarkson, Guedes and Thompson (1996) show that estimation risk of parameters of a security's payoff distribution affect the beta and the expected return of the security. Greater disclosure can reduce such risk and hence the cost of capital. Our results show that information risk (more precisely, price discovery risk) provides another channel through which disclosure can affect the cost of capital.

The remainder of the paper is organized as follows. Section 2 describes data and sample construction. Section 3 introduces VECM and the construction of *ECIN*. Section 4 reports empirical results and Section 5 concludes the paper.

1.2 Data and Sample

The sample includes all ordinary common stocks listed on New York Stock Exchange (NYSE) and American Stock Exchange (Amex) from January 1983 through December 2006. We obtain intraday quote and trade data from Institute for the Study of Security Markets (ISSM) and Trade and Quote (TAQ) databases. For the asset pricing test, we obtain stock returns and characteristics from CRSP daily and monthly files and Compustat annual files. The sample requires stocks to have valid records in

Compustat, CRSP, and ISSM/TAQ. We exclude stocks with less than six months of records in CRSP and stocks with negative or zero shares outstanding. We also delete REIT, stocks of companies incorporated outside the U.S. and closed-end funds. *PIN* data from 1983 to 2001 is obtained from Soeren Hvidkjaer's website. We estimate *PIN* for the following years 2002 to 2006 with the method described in Easley et al. (1996). The ISSM/TAQ intraday data with irregular terms or trades outside regular trading hours are excluded.¹ Open trades are also excluded since those trades are aggregated artificially as large size trades and do not necessarily indicate any private information. To exclude possible reporting errors, we also delete the trade with price change of more than 50% from the previous trade and with reverse price change of more than 50% in the immediately following trade.

We merge ISSM with CRSP by ticker symbols, and TAQ with CRSP by CUSIPs. According to Sadka (2006), this merging method yields the highest matching rate. Moreover, Easley, Kiefer, O'Hara and Paperman (1996) argue that the large trade tends to have multiple participants on one side of the trade. These trades are typically reported as multiple transactions by reporting conventions, while actually only one trade happens. To eliminate this problem, we aggregate all trades occurring within 5 seconds of each other at the same price and with no intervening quote revisions. But

¹ We delete trades and quotes outside the regular trading time 9:30-16:00. For ISSM data, we only include trades and quotes with their condition code as blank or “*”. For TAQ quote data, we exclude quotes with their mode as 4,7,9,11,13,14,15,19,20,27, or 28. For TAQ trade data, we only include trades with the correction indicator as 0 or 1 and the condition code as blank or “*”.

the choice of 5 seconds is somehow arbitrary. To test the robustness, we only aggregate trades occurring within the same second and redo all the following tests. We find this setting does not significantly change any empirical result and all conclusions are robust. Table 1-1 reports sample statistics for the variables in asset-pricing tests.

1.2.1 Large Trades and Small Trades Classification

To construct the information risk measure, we first separate trades into two groups where the most and the least likely informed trades can be found. When informed traders observe private information about the stocks, they will adopt an optimal trading strategy to maximize the trading profits. But at the mean time, they also prevent the private information being quickly revealed through their trades. With these two motivations, informed traders will trade in larger size but not in the extreme ones, since the latter would easily give them away and put them in a disadvantage position. Thus, they may trade through a series of fairly large trades and the private information is revealed gradually. Easley and O'Hara (1987) illustrate in their model that informed traders prefer to trade in large size in order to minimize their trading costs, because every trade incurs a fixed commission fee such that the trading costs from a series of small trades can easily offset the trading profits, especially if the private information is short-lived. Consequently, we believe that the informed trades are mostly likely to be found in the large trade size group and the small trade size group will contain trades that are least likely to be informed.

Barclay and Warner (1993) also find that price movements are mainly due to informed traders' private information and informed trades are concentrated in the medium-size trade category with trade size from 1,000 to 9,900 shares. In their sample, small trades (the trades of below 1,000 shares) account for 52.6% of the total trades but $(-2.3\%)^2$ of the total cumulative price change in the preannouncement period of tender-offer targets, while medium trades account for only 45.7% of the total trades but 92.8% of the cumulative price change. Large trades are defined in their sample as the trades of greater than 10,000 shares. Trades of such a large amount are rare in the exchanges that they only account for 1.7% of total trades and 9.5% of price change. Chakravarty (2001) used a unique audit trail dataset to show further that the source of the large cumulative price impact of medium-size trades is from trades initiated by institutions. As most of these medium trades will be classified as large trades according to our trade size classification scheme described below, this evidence is consistent with our belief that informed trades are mostly likely to be found in the large trade size group.

Following Lee (1992), Lee and Radhakrishna (2000) and Hvidkjaer (2006), we classify the large trades and the small trades according to the trade-size cut-off points that are not only dollar based but also firm size and price specific. The classification method for an individual stock i can be described as follows:

² The negative cumulative change means small trades are so uninformed that all together they change in the wrong sign.

1. In each month t , we sort all stocks into quintiles based on their firm sizes and identify the 99th percentile stock price within each firm-size quintile.
2. For each firm-size quintile, we take the time-series average of 99th percentile stock price over the whole sample period. These averages are denoted as P99.
3. For each firm-size quintile, the cut-off point is $\$100 \times \text{P99}$ rounded up to the nearest \$100.
4. We compare the month $t-1$ closing price for stock i to the dollar-based cut-off points estimated in Step 3 and determine the number of round lot shares that equals to the dollar value of the cut-off point. This number of round lot shares will be the share-based trade-size cut-off point for stock i in month t . Trades with number of share greater (smaller) than the share-based trade-size cut-off points are defined as “large (small) trades.”

This method of classification has the following strengths. First, because the large and small traders should be defined based on their investment value, a dollar-based threshold for classifying trades is most logical. Secondly, this method involves a comparison between the month-end stock price and the dollar-based cut-off points in order to determine the share-based trade-size criterion for the following month, and therefore avoids the sensitivity to small price changes at every single trade as in a simple dollar-based classification criticized by Lee (1992). With the simple dollar-based classification, a \$10,000 small-trade cut-off for a stock traded at \$20 for example would classify all trades of 100-500 shares as small trades. If the price moves to $\$20\frac{1}{8}$ at next trade, only 100-400 shares are classified as small since cut-offs are expressed in round lot. However, investors typically do not change their trading

strategy for every trade in response to one tick's price change. This artificially introduced criterion shifting does not reflect the real trading behavior of investors. Thirdly, the cut-off values are conditional on firm size. Lee and Radhakrishna (2000) illustrate that both large and small traders scale down their trade size when trading in smaller firms, so the classification method conditional on firm size is more accurate. Lee and Radhakrishna (2000) also argue that this classification method maximizes the number of institutional trades in the large trade size group. Given the fact that institutional traders are more likely to be informed, this classification increases the chance of informed trades being correctly classified as large trades.

Since market conditions and trading technologies have changed dramatically in the sample years of 1983 to 2006, the trade-size criterion may be different over time. We therefore divided the sample period into two sub-periods, where the first one is from year 1983 to 1994 and the second is from year 1995 to 2006. The dollar-based trade-size criteria are calculated separately for the two sample periods and reported in Panel A of Table 1-2. We can see the dollar-based cut-offs increases monotonically with firm sizes and the second period observes higher dollar amounts than the first period. Panel B reports the average of the calculated share-based cut-off points for the two sample periods. In particular, we calculate cross-sectional mean of the share-based cut-off points in each month and then take the time-series average of these statistics. The share-based cut-offs are not sensitive to either firm-size or to the sample periods. As previously discussed, we exclude the block trades from the large trades group as those trades tend to be uninformed. The last three columns of Panel B report the average percentage of small, large and block trades in the total number of

trades. Similar to Barclay and Warner (1993), about 50 to 60 percent of trades are in the small trades group and only around 2% are in the block trades group. The remaining trades, which are called medium trades in their paper and large trades in this paper, account for another 40 to 50 percent.³ For robustness check, we also repeat our tests using a simple share-based classification for the whole sample by which trades above 1000 shares are classified as large trades and those below 1000 shares are classified as small trades. The results are similar to those reported in the paper.

We match large and small trades using the MINISPAN procedure outlined in Harris, McInish, Shoesmith and Wood (1995). First we obtain the prices of all large trades for each stock. Then we match each large trade with the small trade nearest in time. Stocks with no large trade or no small trade in a sample year are excluded. In particular, about 1% firm-years are deleted for no matched observations, resulting in 38,594 firm-years in the study. The MINISPAN procedure minimizes the observation time span within each pair by looking forward and backward in the trading time. The average number of MINISPAN matched pairs for one firm-year is 22,328, with year 1984 having the lowest average number at 3,523 and year 2006 having the highest at 93,749. Within each pair, the mean time between the paired large and small trades is 2.06 minutes. For the robustness check, we also match the large and small trades with

³ In the earlier version, we delete about 12% of smaller-sized large trades from the potentially-informed-trades group. This buffer zone between large- and small-trades enhances the large-small classification accuracy, but results in too few observations in the large-trade group for the infrequently traded stocks. And this problem brings bias in the *ECIN* estimation for the infrequently traded stocks, so we add back the buffer-zone stock to the large-trade group.

a restriction that the two trades in one pair must happen within 5 minutes. This method results in a few large trades that are not matched by small trades, but the overall empirical results are insensitive to the imposition of this restriction. In 50.59% of the matched pairs, the large trade precedes the small trade. The average time gap between two pairs is 3 minutes, but this is largely driven by the highly-frequently traded stocks. If we calculate the mean time gap for each firm-year and then take the average, the time gap between two pairs is 17 minutes. Thus we retrieve the matched pairs at fixed intervals of 20 minutes. In every 20 minute interval, we choose the pair closest to and before the interval time point. If there are no trades during an interval, we repeat the last available price pair. The repeated observations for the no-trading periods are 26.34% of the total number of observations in VECM estimations.

1.2.2 Price Impact Analysis of Large and Small Trades

To show that large trades indeed are more likely to be informed trades, we compare the price impact of large and small buy and sell trades as in Holthausen, Leftwich and Mayers (1987, 1990). We use the tick rule to classify trades. If a trade is executed at a price above (below) the previous trade price, it is classified as a buyer-initiated (seller-initiated) trade. If a trade occurs at a price equal to the previous trade, it is ignored. According to Lee and Ready (1991) and Finucane (2000), the tick rule classifies trades with reasonable accuracy. Figure 1.1 plots the cumulative average returns in 20 trades around the seller- and buyer-initiated trades. We can see that large trades have larger price impact but smaller price reversal than small trades on average. There is more

persistence in the price impact of large trades than small trades, which indicates large trades reveal more private information. The positive (negative) return from the trade (-2) to trade (-1) before a down-tick (up-tick) suggests that some down-tick (up-tick) return at trade 0 is caused by the bid-ask bounces and this effect is larger for small trades.

We also analyze the permanent and temporary price effect empirically. Denoting P_s as the logarithm transaction price at time s , we can assume that P_{s-i} is the equilibrium price before the trade s and P_{s+j} is the equilibrium price after the trade s . For example, a seller-initiated trade at time s results in a drop from the previous equilibrium price, so P_s is smaller than P_{s-i} . This price impact may because of information and/or transaction cost. The transaction cost related part will be recovered, while the information related impact will not, so P_{s+j} is greater than or equal to P_s , but smaller than or equal to P_{s-i} . Similarly, for a buyer-initiated trade at time s , we have $P_s \geq P_{s+j} \geq P_{s-i}$. Following Holthausen et al. (1987, 1990), we define $(P_{s+j} - P_{s-i})$ as the permanent price effect of trade s , $(P_s - P_{s+j})$ as the temporary price effect, and $(P_s - P_{s-i})$ as the total price effect. The choice of i and j is somehow arbitrary. Based on the speed of market response, there is no significant price movement fives trades before or after an up or down tick. We accordingly conduct the test with the combinations of i and j choosing from the value of 1 to 5. For brevity, we only report the results with $(i=1, j=1)$ and $(i=3, j=3)$, and the other results are similar.

Table 1-3 reports the calculated permanent, temporary and total price effect for seller-initiated trades in Panel A and for buyer-initiated trades in Panel B. For both large and small trades, the price effects are all negative in seller-initiated trades, while

all positive in buyer-initiated trades. In both panels, we can observe that large trades have a significantly larger permanent price effect than small trades. On the other hand, small trades have higher temporary price effect imitating the strong reversal observed in Figure 1.1. The differences in permanent price effect between large and small trades are significant for both seller-initiated and buyer-initiated trades. In addition, we can see larger differences in the longer time lags ($i=j=3$). This observation suggests that it takes time for market to resume a new equilibrium, and in a shorter time lags ($i=j=1$) trades can be influenced by the market microstructure frictions, e.g. bid-ask bounces.

Although we have confirmed our belief that a large trade is more likely to be an informed trade, not all of the large trades are informed. For the information risk pricing test, we need a measure that allows us to compare cross-sectionally the information risk of potentially informed large trades. *ECIN* is such a measure and we describe its construction in the next section.

1.3 VECM and the Construction of ECIN

As noted previously, *ECIN* is the price discovery of large trades, which can be estimated via the vector error-correction model (VECM). Engle and Granger (1987) provide the cornerstone research linking cointegrated series to the concept of error correction. Cointegration refers to the case in which while multiple series are each non-stationary, they do not diverge without bound from one another. The cointegrating vector defines the long-run equilibrium relationship between these

series, while the error correction describes the short-run dynamics with which each series move towards the long-run equilibrium. For example, two random walk time series X_t and Y_t are each non-stationary and integrated of order one. But if they are cointegrated, they will not diverge from one another without bound. Instead, they will move together through time towards a long run equilibrium relationship described by $Y_t = -\theta X_t$. In this case, the linear combination of X_t and Y_t , defined by $Z_t = Y_t + \theta X_t$, is a stationary $I(0)$ process with zero mean. $\beta = (1, \theta)^T$ is the cointegrating vector and Z_t is the long-run equilibrium error or the extent of disequilibrium. VECM captures the transitional dynamics of the system to long-run equilibrium. In particular, it describe how X_t and Y_t respond to Z_{t-1} , the extent of the disequilibrium in the previous period, in the process of converging toward the long-run equilibrium.

The VECM has been widely used to study the price discovery among prices of same security traded in different markets. For example, Harris, McInish, Shoesmith and Wood (1995) use the VECM to compare the discovery in IBM prices on New York, Pacific, and Midwest Stock Exchanges. Hasbrouck (1995) uses it to study the price discovery of thirty Dow stocks traded on NYSE (dominant market) and regional exchanges (satellite markets). Eun and Sabherwal (2003) use the VECM to investigate the price discovery of cross-listed Canada security in US and Canada markets. Werner and Kleidon (1996) investigate the cointegration of British stocks cross-listed in UK and US markets. Huang (2002) studies the price discovery of quotes in Nasdaq market submitted by the electronic communication networks (ECNs) and by traditional market makers. Booth, Lin, Martikainen and Tse (2002) use VECM to study price discovery of the trades initiated in the upstairs and downstairs markets in Finland. Hendershott and Jones (2005) study the price discovery among Island ECN and the

floor-based exchanges around the time when the Island goes dark. In this paper, we use VECM to investigate the price discovery of large versus small traders. Our method is similar to Harris, McInish, Shoesmith and Wood (1995) and Eun and Sabherwal (2003) where the error-correcting coefficients of price series are used to assess their respective contribution to the price discovery.⁴

We denote the trade price series of large trades and small trades for stock i as $P_{i,t}^L$ and $P_{i,t}^S$ respectively. $P_{i,t}^L$ and $P_{i,t}^S$ are non-stationary but they are cointegrated as the prices of same security. By Granger Representation Theorem in Engle and Granger (1987), if $P_{i,t}^L$ and $P_{i,t}^S$ are cointegrated, there is a VECM representation of the following form

$$\begin{aligned}\Delta P_{i,t}^L &= c_i^L + \alpha_i^L Z_{i,t-1} + \sum_{h=1}^q \phi_{i,h} \Delta P_{i,t-h}^L + \sum_{h=1}^q \delta_{i,h} \Delta P_{i,t-h}^S + \eta_{i,t}^L \\ \Delta P_{i,t}^S &= c_i^S + \alpha_i^S Z_{i,t-1} + \sum_{k=1}^q \lambda_{i,k} \Delta P_{i,t-k}^L + \sum_{k=1}^q \gamma_{i,k} \Delta P_{i,t-k}^S + \eta_{i,t}^S\end{aligned}\tag{1.1}$$

where Δ denotes the first-order time difference (i.e., $\Delta P_{i,t}^L = P_{i,t}^L - P_{i,t-1}^L$ and $\Delta P_{i,t}^S = P_{i,t}^S - P_{i,t-1}^S$), the subscript t represents a 20-minute time interval, c_i^L and c_i^S are constants, $Z_{i,t-1} = P_{i,t-1}^L - \theta_i P_{i,t-1}^S$ is the equilibrium error term, a stationary process

⁴ We choose the error-correcting adjustment parameters over Hasbrouck (1995)'s information share as the measure for assessing the price discovery because of the same consideration as discussed in Eun and Sabherwal (2003) – the non-uniqueness of the information share measure complicates our cross-sectional analyses that examine the price impact of the information risk.

with mean zero, and $(1, \theta_i)'$ is called the cointegration vector. As the prices of same stock, $P_{i,t}^L$ and $P_{i,t}^S$ should be equal in equilibrium. We have $(\theta_i = -1)$ and the equilibrium error $Z_{i,t-1}$ becomes $(Z_{i,t-1} = P_{i,t-1}^L - P_{i,t-1}^S)$. α_i^L and α_i^S are the error-correcting coefficients that indicate the extent to which the price series of large trades and small trades respond to the disequilibrium (or equilibrium error) during the process of moving toward the long-run equilibrium relationship $P_{i,t}^L = P_{i,t}^S$. For example, if the disequilibrium is positive (i.e., $P_{i,t-1}^L > P_{i,t-1}^S$), large trade price was above the small trade price in the previous period. To maintain the long-run equilibrium, this gap has to be reduced by lowering the large trade price by the amount of α_i^L times the level of disequilibrium $(P_{i,t-1}^L - P_{i,t-1}^S)$, and raising the small trade price by the amount of α_i^S times the level of disequilibrium $(P_{i,t-1}^L - P_{i,t-1}^S)$. This implies α_i^L is zero or negative while α_i^S is zero or positive. In this paper we use α_i^L and α_i^S to assess the price discovery of large and small trades respectively.

The idea that error-correcting coefficients in VECM measure the price discovery can be illustrated with a simple example analyzed in Hasbrouck (1995)

$$\begin{aligned} P_t^1 &= P_{t-1}^1 + w_t \\ P_t^2 &= P_{t-2}^1 + \varepsilon_t \end{aligned} \tag{1.2}$$

where P_t^1 and P_t^2 are prices of same security traded by two group of investors (or in two different markets). w_t and ε_t are zero-mean uncorrelated disturbances that are identically and independently distributed. The price traded by the first group (P_t^1) follows a random-walk with increment w_t . The prices traded by the second group (P_t^2) tracks the first price lagged by two periods and also reflects a random error possibly caused by some random liquidity demand. It is clear in this example that all of the

random-walk variation arises from the first price and hence that the entire price discovery takes place in the first price.

It is straightforward to show that

$$\begin{aligned} P_t^1 &= P_0^1 + \sum_{s=1}^t w_s \\ P_t^2 &= P_0^2 + \sum_{s=1}^t w_s + (-w_t - w_{t-1} + \varepsilon_t) \end{aligned} \quad (1.3)$$

where P_0^1 and P_0^2 are the initial values of the first price and the second price respectively. $\sum_{s=1}^t w_s$ can be viewed as the implicit efficient price which is also the common trend shared by P_t^1 and P_t^2 that makes the two price series cointegrated (Stock and Watson (1988)). By combining P_t^1 and P_t^2 according to $I_t = P_t^1 - P_t^2$, the common trend is removed and I_t becomes a stationary $I(0)$ process with zero mean (assuming $P_0^1 = P_0^2$). In the long run, both P_t^1 and P_t^2 converge to the efficient price so that $P_t^1 = P_t^2$ (again assuming $P_0^1 = P_0^2$) defines the long-run equilibrium and I_t can be viewed as the level of disequilibrium or long-run equilibrium error. As we can see from equation (1.3), P_t^1 is always at the long-run equilibrium price because it reflects the efficient price perfectly all the time.

As P_t^1 and P_t^2 are cointegrated, it can be easily verified that the following VECM representation exists.

$$\begin{aligned} \Delta P_t^1 &= w_t \\ \Delta P_t^2 &= (P_{t-1}^1 - P_{t-1}^2) - \Delta P_{t-1}^1 + \varepsilon_t \end{aligned} \quad (1.4)$$

If we let the first price be the large trade price (i.e., $P^1 = P_i^L$) and let the second price be the small trade price (i.e., $P^2 = P_i^S$), comparing equation (1.1) and (1.4) reveals among others that $Z_{i,t-1} = (P_{t-1}^1 - P_{t-1}^2)$, $\alpha_i^L = 0$ and $\alpha_i^S = 1$. In this example, the large trade price is always the long-run equilibrium price, and is therefore responsible for all the price discovery, so we expect it not to react to the disequilibrium in the previous period, $(P_{t-1}^1 - P_{t-1}^2)$. This expectation is captured by $\alpha_i^L = 0$. In contrast, the small trade price has no price discovery function and the disequilibrium in the previous period was caused completely by its deviation from the long-run equilibrium price. Thus, we expect the small trade price to respond fully to the disequilibrium by eliminating the said deviation. This observation is consistent with $\alpha_i^S = 1$. In other words, all of the adjustments toward the long-run equilibrium fall on the small trade price, and the large trade price does not respond to the disequilibrium. Thus zero adjustment of the large trade price ($P_i^L = P^1$, $\alpha_i^L = 0$) indicates that the entire price discovery occurs in the large trade. Consequently, all of the adjustment towards the long-run equilibrium falls on the small trade price, while the large trade price on the contrary does not respond to the disequilibrium.

On the other hand, the disequilibrium can be purely because of the trading friction. Price drops for a seller-initiated trade, while increases for a buyer-initiated trade. Since there is no information involved, all price impact will get recovered in the future trades. In that case, we observe the trade which initiates the disequilibrium responds to the disequilibrium. Because large trades typically have large price impact and reversal if the stock is illiquid, α_i^L is negative with large magnitude. α_i^S in this case also assess the trading friction of small trades.

In more reality cases, the disequilibrium is a mixed effect of both information and trading friction. When large trades initiate a buy-order, for example, the price increase may have two components; one reveals favorite information and the other is the extra trading cost. In the future trades, the part related to the information persists but the part related to the trading cost reverses. The larger of the persistence, the more information components is in the price impact, and therefore the larger α_i^L (closer to zero) is. On the other hand, if further price reverse a lot, there is not much price discovery and we observe small α_i^L . Thus, the same intuition as in the simple example holds: if the price of a trade is less responsive to the level of disequilibrium (smaller error correcting coefficient in absolute value), it would be closer to or reflect more of the efficient price and therefore smaller adjustment is needed in this price to maintain the long-run equilibrium. Moreover, if informed trader purposely split their total trade into several large trades to hide their information, we may observe a positive α_i^L , suggesting a price momentum in large trades. And this price momentum also indicates that the probability of informed trading is high.

We run the VECM every firm-year. In a period of one year, a firm does not necessarily have only informed trades or uninformed trades in the large trades group. In most cases, there are both informed and uninformed large trades. The fraction of informed trades depends on both the frequency of information events and the private signal observed by informed trades. A higher fraction of informed trades in the total number of large trades similarly leads to more price continuity in large trades and consequently larger α_i^L for that firm-year.

According to the definition of price discovery, large trades whose prices reflect more of the efficient prices are more likely to be information-based trades. We compare $|\alpha^L|$ or α^L across firms. The firm with larger α^L or smaller $|\alpha^L|$ ($\alpha^L < 0$) has more price discovery in the large-sized trades. Given that informed traders prefer to trade in large size, more price discovery in the large-size trades suggests more intensive informed trading in this stock. Consequently, α^L serves as a natural measure to assess the intensity of informed trading or the information risk, which we have also referred to as *ECIN* in this paper.

Based on the discussion above, we present the following hypotheses:

- 1) α^L is positively related to the degree of information asymmetry.
- 2) If information risk is a priced risk, α^L will be positively related to the future stock returns.

1.4 Empirical Results

The cointegration and error-correction analysis is conducted in each calendar year for each firm. We estimate the model with the lag q equals 5, 10 and 15, separately. The main parameters and test results do not vary significantly with these settings, so we only report those with q equals 5 in the paper. Since the time gap between P_t and P_{t+1} is 20 minutes, a model with q equals 5 controls for the autocorrelation in lagged price changes up to 100 minutes. We first conduct the augmented Dickey-Fuller unit-root test to check if both P^L and P^S are $I(1)$. Then we perform the Johansen (1988)

cointegration test to check whether P^L and P^S are cointegrated.⁵ We start with a sample of 38,594 firm-year observations. In total, 3,368 firm-year observations fail either the unit-root test or the cointegration test and are deleted. We have 35,226 firm-year observations in our final sample.

Panel A of Table 1-4 reports the summary statistics of VECM parameters by pooling all firm-years. The median of α^L is (-0.172) and the median of α^S is 0.182. We also report the standard deviation σ of the error terms η_t^L and η_t^S , which is 0.123 and 0.121 respectively for large and small trades. The slightly smaller $|\alpha|$ and higher σ in medians for large trades suggest that at least 50% sample firm-years have more price discovery in large trades than in small trades. In total, we have 56.6% of firm-years with negative $(|\alpha^L| - \alpha^S)$, which also means that we have 43.4% of firm-years with $|\alpha^L|$ larger than α^S . This is not surprising given that the quantity of private information differs across firms. In some firms, the large trades are as uninformed as the small trades are, and the stock price deviates from the equilibrium mainly due to the illiquidity related transaction cost. In fact, a large dispersion in α^L would enhance the power of empirical tests.

Panel B reports the characteristics of $(P_i^L - P_i^S)$, ΔP_i^L , and ΔP_i^S . We first calculate the mean and standard deviation of these variables for each firm-year and

⁵ The unit-root test and cointegration test are conducted without the trend term. The tests with the trend term result in a slightly different sample of firm-year observations that fail the unit-root of cointegrations test, but all the empirical results and conclusions remain unchanged.

then report the median and percentiles of these means and standard deviations. Panel B also reports the correlation of α^L with the means or standard deviations of these parameters. The median of the average ΔP_i^S is smaller than that of the average ΔP_i^L , suggesting that large trades initiate more price changes. But the range (P90-P10) of average ΔP_i^S is much larger than that of average ΔP_i^L . In addition, the standard deviation of ΔP_i^S in each firm-year is also larger than that of ΔP_i^L . This suggests that small trades are noisier than large trades. α^L is negative correlated with all the average price changes $(P_i^L - P_i^S)$, ΔP_i^L , and ΔP_i^S . When trading friction or transaction cost is large, stocks have more disequilibrium (differences in large and small trades) and all price changes should be large on average but do not reflect information. On the other hand, the standard deviation of $(P_i^L - P_i^S)$ is positively correlated with α^L . If the difference in prices of large and small trades is caused by the private information, the level of the difference largely depends on the magnitude and accuracy of private information, which varies from time to time. But if the price difference is mostly transaction cost, it should be around the same level and have smaller variation. Similarly, if most responds of small price changes is the price impact or reversal caused by transaction cost, it would have small variation. But if it is the responds to the price discovery in large trade, the changes will depend on different private information and be more volatile. Therefore the standard deviation of ΔP_i^L is also positively correlated with α^L . On the other hand, the standard deviation of ΔP_i^L reflects how the prices of large trades respond to the disequilibrium. A large and more volatile change in large prices suggests that the large trades respond more to the disequilibrium and therefore smaller price discovery.

Panel C reports the summary statistics of *ECIN* measure by year. We can observe *ECIN* decreases significantly over time. This trend can be driven by the trading innovations (e.g. decimalization, high-frequency and algorithmic trading), better regulations (e.g. Regulation Fair Disclosure and Sarbanes–Oxley Act), or booming financial markets with much more uninformed traders. It is an interesting future topic to study how much these factors affect the information asymmetry and information-based trades.

1.4.1 The Relationship of ECIN with Traditional Information Asymmetry Measures

Strictly speaking, hypothesis (1) can only be jointly tested with hypothesis (2). However, if *ECIN* is a good information risk or information asymmetry measure, we would expect it to be positively related to the traditional information asymmetry measures. Some researchers (e.g. Lee, Mucklow and Ready (1993)) argue that firms with higher bid-ask spreads have higher information asymmetry. Lo and MacKinlay (1990) argue that large firms have a lower degree of information asymmetry or smaller adverse selection costs. From Panel A of Table 1-5, we see *ECIN* is positively correlated with bid-ask spread and negatively correlated with firm size. We also examine the link of *ECIN* to *PIN*. Developed by Easley, Kiefer, O'Hara and Paperman (1996), *PIN* has become one of the most popular measures of information risk, measuring the probability of private information-based trades. Panel A reveals that *ECIN* and *PIN* are positively correlated at 0.484. The correlation is modest,

suggesting the information risk test results based on *PIN* may not carry over to those based on *ECIN*.

Panel B of Table 1-5 reports the value of traditional information asymmetric measures in each quintile sorted by *ECIN*. Quintile five has the highest *ECIN* and quintile one has the lowest *ECIN*. The sorting is done each year and the figures reported are the yearly average of firm size, bid-ask spread, and *PIN* in each quintile. We can see clearly that the size is decreasing in *ECIN*, while the bid-ask spread and *PIN* are all increasing in *ECIN*. In sum, *ECIN* possesses the properties that are expected of good information risk measure.

1.4.2 Asset Pricing Test

In this section, we perform the major task of this paper: a joint test of hypothesis (1) and (2) that *ECIN* is an information risk measure and that information risk is priced. We estimate the Fama and MacBeth (1973) (FM thereafter) regressions with Newey-West standard errors. First the cross-sectional regression is estimated for each month in the sample period over year 1984 to 2007. Then the coefficients of cross-sectional regressions are averaged through time. And *t*-statistics of the time-series means are based on Newey and West (1987) autocorrelation- and heteroskedasticity-consistent standard errors. Like Easley et al. (2002), we also employ Litzenberger and Ramaswamy (1979) weighted least-square (LR-WLS thereafter) methodology to correct the time-varying volatility, which involves weighting the coefficients of each month's cross-sectional regression by the ratio of the precision of the estimation in that month to the average precision through the whole period. The dependent variable is

the month t return (in percentage) in excess of the one-month T-bill rate. The monthly return data are from the Center for Research in Security Prices (CRSP) files. $ECIN$ is the error-correcting coefficient estimated in previous calendar year from equation (1.1). Beta is the post-ranking portfolio betas of Fama-French (1992) 40 post-ranking portfolios that are constructed based on the pre-ranking betas of individual stocks which are estimated using monthly returns from at least two years to, when possible, five years before the test year. Firm size ($LN MV$) is the logarithm of the market value of equity at the previous year end, book-to-market ($LN BM$) is the logarithm of the book value divided by the market value at the previous year end, and $R6$ is the six-month cumulative stock return from month $t-2$ to $t-7$. Book values of equity are from Compustat annual files (item 60). Firms with negative book values are excluded. And BM is winsorized at the 0.5% and 99.5% levels.

Table 1-6 reports both FM and LR-WLS coefficients of the asset pricing tests. Regression (1) serves as a check of our data and methodology by regressing returns on some variables that have been known to predict returns—the lagged year's beta, firm size and book-to-market ratio. Consistent with prior findings, Regression (1) shows that stocks returns are positively related to BM and that there is an insignificant size effect in the post-80 period. It also shows that returns are negatively related to beta. Although inconsistent with CAPM, this result is also found in Fama and French (1992), Chalmers and Kadlec (1998), Datar et al. (1998) and Easley et al. (2002) that investigate similar sample periods.

Regression (2) reports our baseline result results for which our key variable, the lagged year $ECIN$, is included in the regression in addition to the independent

variables we have seen in Regression (1) as controlling variables. The coefficients of *ECIN* are significantly positive with *t*-statistic of 3.47 in FM regression and a *t*-statistic of 3.30 in LR-WLS regression. This result supports the joint hypothesis (1) and hypothesis (2) that *ECIN* captures information risk (i.e., it is positively related to the degree of information asymmetry) and it is priced in the equity market. In addition, when *ECIN* is controlled, firm size has a significantly positive coefficient. This suggests that the higher return in the smaller firms captured by the information risk. Consistent with the decreasing information risk in recent years (Panel C of Table 1-4), the “size effect” is disappearing or even becomes positive in the more recent periods. We then check whether the pricing effect of *ECIN* can be explained by the well-documented momentum effect (Jegadeesh and Titman (1993)) in regression (3) by adding the past six-month return (*R6*) as an additional control variable. *ECIN* remains highly significant, suggesting that the pricing effect of *ECIN* is not related to the price momentum.

Duarte and Young (2009) argue that *PIN* is priced because it captures the effect of illiquidity. When an illiquidity measure such as Amihud’s (2002) *ILLIQ* is controlled for in the asset pricing test, the significance of *PIN* diminishes. To make sure this is not the case with the pricing impact of *ECIN*, we also control for *ILLIQ* in our tests. Following Amihud (2002), *ILLIQ* is calculated as

$$ILLIQ_{iy} = \frac{1}{D_{iy}} \sum_{t=1}^D \frac{|R_{iyd}|}{VOLD_{iyd}} \quad (1.5)$$

where D_{iy} is the number of days for which data are available for stock i in year y , R_{iyd} is the return on stock i on day d of year y and $VOLD_{iyd}$ is the corresponding daily volume in dollars.

The positive and significant coefficient on *ILLIQ* in regression (4) shows that the liquidity effect that Amihud (2002) documented is also present in our sample and may potentially explain the pricing impact of *ECIN*. However, in Regression (5) when both *ECIN* and *ILLIQ* are included in the independent variables, we find that rather than explaining the pricing impact of *ECIN*, the liquidity effect of *ILLIQ* itself changes sign and is subsumed by *ECIN*. This result is interesting because it shows that the liquidity effect that *ILLIQ* captures originates from information asymmetry. This is consistent with the theoretical models of Kyle (1985), Glosten (1989) and Easley and O'Hara (1987). Brennan and Subrahmanyam (1996) who argue that the primary cause of illiquidity in the financial market is the adverse selection due to the presence of informed traders. They further argue that the liquidity effects of the asymmetric information are most likely to be captured in the price impact of a trade, or the variable component of trading cost that *ILLIQ* is designed to capture.

We conduct further checks to make sure the pricing impact of *ECIN* is not driven by other well-known effects or affected by other measures of liquidity. For brevity, we only report the FM results of these tests in Table 1-7⁶.

We first check if the pricing impact of *ECIN* is driven by the tax-loss selling induced January effect (cf. Rozeff (1976), Roll (1983), D'Mello, Ferris and Hwang (2003), Grinblatt and Moskowitz (2004).) This concern arises from the fact that high *ECIN* firms also tend to be small firms as revealed earlier in Table 1-5. Research in tax-loss selling and January effect literature has documented a large January return for

⁶ The unreported LR-WLS regression results are similar.

small firms; this may explain the large return of high *ECIN* firms. Although we have controlled for firm size in our regressions, to fully address this concern we exclude January return when we run Regression (1) in Table 1-7 and we can see that *ECIN* remains positively significant. Then we run the regression on Januarys only and report the results in Regression (2). We can see *ECIN* is negatively priced in Januarys. That means high (low) *ECIN* firms tend to have low (high) January effect. Because the tax-loss selling induced big positive return in January especially for illiquid firms, the negative coefficient on *ECIN* in Januarys is evidence that *ECIN* are not capturing the illiquidity. In Regression (3) to (10), we test whether other variables can explain the pricing effect of *ECIN*. Regression (3), (5), (7) and (9) are run in non-January months, while Regression (4), (6), (8) and (10) are run in Januarys.

We first check if the pricing impact of *ECIN* can be explained by turnover. A low turnover means that traders cannot assume or unload positions easily without incurring a large transaction cost, thus we expect low turnover stocks to have high required returns. Consistent with the illiquidity risk premium hypothesis, Chordia, Subrahmanyam and Anshuman (2001) find that turnover is indeed negatively related to the future return. We follow Chordia, Subrahmanyam and Anshuman (2001) to include the coefficient of variation of turnover (*CVT*) along with the share turnover (*TURN*) in the asset pricing test. *TURN* is the natural logarithm of the share turnover measured by the number of shares trading volume divided by the number of shares outstanding in the second to last month. *CVT* is the natural logarithm of the coefficient of variation of turnover calculated over the 36 months period prior to second to last month. The result is reported as Regression (3) and (4) in Table 1-7. *ECIN* is positively significant in non-January months and negative in Januarys. We also check whether the pricing impact of *ECIN* is related to bid-ask spread. High bid-

ask spread firms are thought to be less liquid due to the high transaction costs related to spread. Amihud and Mendelson (1986) find a positive relationship between return and bid-ask spread which is again consistent with illiquidity risk premium hypothesis discussed above. Bid-ask spread variable, *SPR*, is measured as the average of daily opening percentage spreads in the previous year. The testing result that controls for *SPR* is reported in Regression (5) and (6). In Januarys, *SPR* is positively significant, suggesting illiquid firms have higher January returns. But *ECIN* is still negative in January, indicating *ECIN* is not capturing the illiquidity as *SPR* does. In non-January months, *ECIN* is positively significant, while *SPR* is not. Note that bid-ask spread is also often used as the proxy for information risk since it contains an adverse selection component (c.f. Glosten and Harris (1988), George, Kaul and Nimalendran (1991), Huang and Stoll (1997), Lin, Sanger and Booth (1995), and Madhavan, Richardson and Roomans (1997)). Our result indicates *ECIN* is a better information risk measure than bid-ask spread to test the pricing impact of information asymmetry. Finally, we check whether *ECIN* is a proxy for return volatility. Liu and Wysocki (2006) argue that information risk does not affect the cost of capital after controlling for the standard deviation of returns. We define *STD* as the standard deviation of daily returns in year $t-1$. The results are reported as Regression (7) and (8). The negative coefficient on *STD* in Regression (7) shows that the volatility has a negative effect on the stock return, which is consistent with Ang, Hordrick, Xing and Zhang (2006) and *ECIN* is still significant. In January, *STD* is positive, while *ECIN* is negative. In Regression (9) and (10) of Table 1-7, we include all the variables discussed above to check if the pricing effect of *ECIN* can be jointly explained by these variables. The

result indicates that *ECIN* remains significantly priced after controlling for all these variables.

Since *ECIN* is an estimated variable, it is potentially subject to an errors-in-variables (EIV) bias. In the asset pricing tests, beta is another estimated variable, but the Fama and French (1993) portfolio approach is to correct for this problem. If we assume the EIV problem has been corrected for beta, the only variable in the problem is *ECIN*. According to Wooldridge (2001) and Levi (1973), if only one variable in the regression is subject to the EIV bias, the coefficient on this variable is biased down, and therefore the bias is against finding any effect on *ECIN*. Nevertheless, we can also check the importance of this problem by conducting an instrument test following Easley, Hvidkjaer and O'Hara (2002). Similar to their method, we sort forty portfolios in each year, $t-2$, according to the estimated *ECIN*. The portfolio *ECIN* is the average of the estimated *ECIN* in year $t-1$. We assign the portfolio *ECIN* to each stock in the portfolio and run the asset pricing test with this instrument. The portfolio *ECIN*, denoted as *PECIN*, is correlated with *ECIN* at 0.655. In addition, we can argue this instrument is not correlated with the error term. If there is an EIV problem, then the instrument will fix the problem by driving down the coefficient of *ECIN* to zero. But if there is no EIV problem, the coefficient on the instrument should be zero. We include both *PECIN* and *ECIN* in the regression. The results in regression (7) of Table 1-7 illustrate that the coefficient of *ECIN* is still significant, and that the coefficient on *PECIN* is not significantly different from zero. Thus, we can conclude that the portfolio approach does not correct for any EIV problem, and we should focus on the original *ECIN* rather than the instrument *PECIN*.

In the estimation of *ECIN*, we retrieve large and small trade pairs in every 20 minutes interval. When there is no such trade pair in a particular interval, we will repeat the most recent available trade pair. Those repeated observations may result in an overestimated *ECIN*, since the price of the later large trade is same as the price of the previous large trade. We estimate the percentage of the number of the repeated observations over the total number of all observations for each firm-year. On average, 26.34% observations are repeated price-pairs from the previous trades. We control this percentage in the asset pricing regression to test. If the pricing effect of *ECIN* is driven by the infrequent trading, we will observe the percentage of the repeated observations subsumes *ECIN*. The results in regression (8) of Table 1-7 illustrate that the coefficient of *ECIN* is still significant controlling for the percentage of repeated observations.

In sum, we have shown that information risk measured by *ECIN* is priced. Furthermore, the pricing effect of *ECIN* survives the control of a group of variables and effects that have been known to affect returns. In particular, it survives the control of liquidity related variables individually or jointly. This is in sharp contrast to the pricing effect of *PIN* which Duarte and Young (2007) have shown to be caused by the liquidity effect.

The asymmetry in robustness of the pricing impact of *PIN* and *ECIN* with respect to the control of illiquidity measures is not surprising. The advantage of *ECIN* is that it is inferred from price discovery directly whereas *PIN* is inferred from abnormal order flow imbalance. Although informed trades generate some abnormal order imbalance, Duarte and Young (2009) argue that such abnormal order imbalance

can also be generated by illiquidity, thus high *PIN* can indicate high illiquidity as much as high information risk. In contrast, *ECIN* assesses the possibility and the significance of informed trades directly from the trading prices and is unlikely to be positively correlated with illiquidity for the reasons explained below.

According to Llorente, Michaely, Saar and Wang (2002), a trade price is likely to continue if the trade is driven by information; this is particularly true when the trade is accompanied by heavy trading volume. In contrast, if a trade is liquidity-motivated; the trade price will reverse and the reversal is especially large for the illiquid stocks. As the heavy trading volume are typically accompanied by large-size trades, the price continuity associated with the informed trade will result in a large price discovery in large trades in VECM and therefore a high *ECIN*. On the contrary, the illiquidity-induced price reversal associated with the liquidity-motivated large trades will result in a small price discovery and hence a low *ECIN*. Furthermore, Hwang and Qian (2009) construct an information risk factor (*ECINF*) based on *ECIN* and show that it is capable of explaining a few well-known market anomalies including price momentum, the post-earning announcement drift (PEAD) and the long-term underperformance of stocks that have undertaken secondary equity offerings. In particular, they show that loser stocks in price momentum and stocks with large negative earnings surprise in PEAD have significantly lower loadings on *ECINF*. These results are consistent with the hypothesis that *ECINF* is an information risk factor and that there are fewer informed traders in losers stocks and stocks with negative earnings surprise. These stocks have fewer informed traders because short selling is required for informed traders to take advantage of the bad news and that short selling can be difficult and risky. On the other hand, if *ECIN* captures illiquidity rather than information risk, then *ECINF* becomes an illiquidity factor; lower loadings on *ECINF* of these stocks implies

that loser stocks or stocks with extreme negative earnings are more liquid than other stocks, which is counterintuitive and hard to fathom. Consequently, a high *ECIN* is unlikely to capture high illiquidity, dissimilarly to *PIN*. This is exactly the reasoning that has led us to conclude earlier that the liquidity effect captured by *ILLIQ* and subsumed by *ECIN* has its origin in information asymmetry.

Note that the pricing effect of our information risk measure, *ECIN*, survives the control, individually or jointly, of a group of major liquidity measures in the literature. This offer powerful support to O'Hara (2003) and Easley and O'Hara (2004). They argue that although the source of price discovery risk and some of the liquidity measures such as bid-ask spread and *ILLIQ* can be traced back to asymmetric information, the pricing effect of the price discovery risk is distinct from that of liquidity. Price discovery risk affects the cost of capital through its effect on the risk faced by uninformed traders who are unable to incorporate new information into their portfolio decisions as informed investors are. In contrast, liquidity affects the cost of capital through its effect on the transaction costs introduced by the market makers. The liquidity effect of asymmetric information can exist in the world of risk-neutral traders. On the other hand, the effect of the price discovery risk is present only when traders are risk adverse.

1.4.2.1 Compare ECIN with PIN

Although the MLE estimation of *PIN* faces convergence problems in the recent sample period, it remains the most popular choice of the information risk measure. Thus, it is

interesting and important to compare the pricing impact of *ECIN* and *PIN*. We run a horse race between the two by adding *ECIN* in the asset pricing test. Again, for brevity we only report the FM coefficients of the tests in Table 1-8 as the results based on the LR-WLS coefficients are similar.

Regression (1) is run on *PIN* controlling for beta, firm size and book-to-market ratio. We can observe *PIN* has a significant positive coefficient, consistent with results in Easley et al. (2002). But comparing Regression (1) in Table 1-8 and Regression (2) in Table 1-6, *ECIN* has higher significance (t -statistics=3.47) than *PIN* does (t -statistics=1.67). Furthermore, when *ECIN* and *PIN* are included jointly in Regression (2), *PIN* becomes insignificant, while *ECIN* remains significantly positive. In Regression (3) and (4), we run the test with Amihud (2002) illiquidity measure (*ILLIQ*). Similar to Duarte and Young (2009), we find *ILLIQ* subsumes *PIN*. When putting together, *ILLIQ* is significantly positive but *PIN* is not significant. When *ECIN* is included in the test, neither *PIN* nor *ILLIQ* is significant, suggesting *ECIN* captures information risk better and is not subsume by the illiquidity measure.

1.4.2.2 Risk Adjusted Return and Economic Significance

Although the positive significant coefficients on *ECIN* in Table 1-6, Table 1-7 and Table 1-8 show that information risk is priced, it is not straightforward to gauge the economic significance of the pricing impact from them. Furthermore, those results are not adjusted for factor risks. In this section, we use the regression approach (cf. Grinblatt and Moskowitz (2004), George and Hwang (2004, 2007)) to form the portfolios of high and low information risk that hedge out the influence of beta

(*BETA*), firm size (*LNMV*), book-to-market (*LNBM*) and the price momentum (*R6*). The economic significance of the information risk can be assessed based on the returns of these portfolios which can be read immediately from the coefficients of the regression. The risk adjusted return can also be easily obtained by further running time series regressions of these coefficients on the contemporaneous risk factors. Specifically, we first run the following cross-sectional regression.

$$R_{i,t} = \gamma_{0,t} + \gamma_{1,t}HECIN_{i,t-1} + \gamma_{2,t}LECIN_{i,t-1} + \gamma_{3,t}BETA_{i,t-1} + \gamma_{4,t}LNMV_{i,t-1} + \gamma_{5,t}LNBM_{i,t-1} + \gamma_{6,t}R6_{i,t-1} + e_{i,t} \quad (1.6)$$

In each month, stocks are sorted into quintiles according to *ECIN* obtained in the previous year. We then define *HECIN* (*LECIN*) dummy that takes the value of 1 for stocks belonging to the top (bottom) *ECIN* quintile and takes the value of 0 otherwise. We further demean *BETA*, *LNMV*, *LNBM* and *R6* by deducting the cross-sectional average of these variables from the values of the respective variables. According to Fama (1976), $\gamma_{0,t}$ can be interpreted as the return in month t of the medium (middle three quintiles) *ECIN* portfolio with the average beta, size, book-to-market and momentum. In other words, it is the return of a medium *ECIN* portfolio that has hedged out the effect of beta, firm size, book-to-market and momentum (benchmark portfolio). Similarly, $\gamma_{1,t}$ ($\gamma_{2,t}$) is the excess return, relative to the benchmark portfolio, of the *HECIN* (*LECIN*) portfolio that has hedged out size, book-to-market and momentum effect. The results are reported in Panel A of Table 1-9. The FM coefficients $\gamma_{0,t}$, $\gamma_{1,t}$, and $\gamma_{2,t}$ are 1.130%, 0.030% and (-0.142%)

respectively. This means the returns of low, medium and high information risk portfolios that have hedged out the size, book-to-market and momentum effect are 0.988%, 1.130% and 1.160% per month respectively. This also means that the return premium of information risk (HL) that has hedged out the effect of beta, firm size, book-to-market and momentum, measured as the return difference between $HECIN$ and $LECIN$ portfolios is 0.172% per month. A similar interpretation applies to the LR-WLS coefficients and the return premium of information risk (HL_{LR-WLS}) is 0.266% per month.

We further test if HL_{FM} and HL_{LR-WLS} can be explained by the popular risk factors, and equivalently test to see if the zero investment portfolio that takes a long position in the $HECIN$ portfolio and a short position in the $LECIN$ portfolio generates a significant risk-adjusted return. This is performed by running the time-series regressions of HL_{FM} and HL_{LR-WLS} on the contemporaneous Fama and French (1993) three factors (MKT , SMB and HML) and Carhart (1997) momentum factor (UMD). The risk-adjusted returns are the intercepts from these regressions; the results are reported in Panel B of Table 1-9. The risk-adjusted return of HL_{FM} and HL_{LR-WLS} are 0.290% and 0.385% per month respectively. Both are economically and statistically significant (with t-statistics=2.63 and 4.25 respectively), suggesting that the risk premium of $ECIN$ cannot be explained by the standard risk factors.

We also examine whether the liquidity factors of Pástor and Stambaugh (2003) and Sadka (2006), which have demonstrated some success in asset pricing, can explain the risk premium or the pricing impact of $ECIN$. We do this by adding liquidity factor as an additional factor in the time-series regressions.

Pástor and Stambaugh (2003) construct non-traded liquidity factors based on the idea that order flow, the volume signed by the contemporaneous return in excess of the market return, should be accompanied by a large return reversal for illiquid stocks. They further construct traded liquidity factors as the returns of the top-decile liquidity beta portfolio minus the returns of the bottom-decile liquidity beta portfolio. The liquidity beta is the stock returns' sensitivity to the non-traded liquidity factors. The equally-weighted and value-weighted Pástor and Stambaugh (2003) traded liquidity factors are obtained from WRDS.

Sadka (2006) utilizes transaction data to construct liquidity factors that aims at capturing the variable component of trades' price impact. This is similar in concept to Amihud's illiquidity measure, but Amihud (2002) uses daily data and focuses on the level of liquidity. We follow the same approach as in Pastor and Stambuaugh (2003) to construct the Sadka's traded liquidity factor from his non-traded liquidity factor. At the end of each year, we estimate Sadka's liquidity beta for each stock by regressing its historical monthly return on Sadka's non-traded liquidity factor together with the Fama-French and Carhart factors. We require that there are at least four years of historical returns for the regression. We then sort the stocks into deciles according to the liquidity beta. The differences of the returns of the top and bottom deciles portfolio for the next 12 months are the factor returns. We repeat the procedure at the end of every year and string all the monthly returns together to obtain the traded Sadka liquidity factor. Both the equal-weighted and value-weighted portfolios returns are calculated. The non-traded Sadka factor is from WRDS and it began in 1983, thus the

traded factors began in 1987 with the four-year minimum data requirement for the liquidity beta estimation.

The liquidity risk adjusted results are reported in Panel B of Table 1-9. We only report the results with the value-weighted liquidity factors as the results with the equally-weighted liquidity factors are quite similar. The loadings of HL_{FM} and HL_{LR-WLS} portfolios on the liquidity factors are not significant and the risk-adjust return of HL_{FM} and HL_{LR-WLS} are still significantly at 0.29 to 0.38%. These results suggest that the return premium of $ECIN$ is either unrelated to or cannot be explained by liquidity risk factors.

1.5 Conclusion

We find that large trades have a larger permanent price effect than small trades have. The price impacts of large seller-initiated and buyer-initiated trades do not reverse in the following trades, but the price impacts of small trades do. These results support the argument of Easley and O'Hara (1987) and Barclay and Warner (1993) that large trades contain more information than small trades. These results also provide the empirical grounding for the $ECIN$ measure that is based on the price discovery of large trades, estimated from the vector error correction model. $ECIN$ possesses the properties that one expects to find in a good information risk measure: it decreases in firm size and trading volume, increases in bid-ask spread and return standard deviation, and is positively related to PIN .

We contribute to the literature by showing that information risk is indeed priced in light of the contradictory evidence from tests based on PIN (Easley, Hvidkjaer and O'Hara (2002) and Duarte and Young (2009).) $ECIN$ is shown to have

a significant positive impact on monthly returns after controlling for firm size, book-to-market ratio, and momentum. A portfolio that takes a long position in the top *ECIN* quintile stocks and a short position in the bottom *ECIN* quintile stock generates risk-adjusted returns of 0.29% to 0.39% per month which is both statistically and economically significant. We also find that the pricing impact of *PIN* is subsumed by *ECIN*. Furthermore, in contrast to *PIN*, the pricing effect of *ECIN* is robust to various illiquidity measures. These results are consistent with the argument of O'Hara (2003) and Easley and O'Hara (2004) that risk of price discovery matters beyond the extent that it affects the transaction costs that confront investors. We also find the result that the pricing impact of *ILLIQ* is subsumed by *ECIN* particular interesting as it suggests that the liquidity effect captured by *ILLIQ* has its origin in information asymmetry.

Since its inception in 2002, *PIN* has received wide acceptance and has been extensively used as an information risk measure in the literature. In this paper, we show that *ECIN* has advantages over *PIN* not only because it is easy to estimate, but also because it is devoid of the convergence problem that can plague *PIN* estimation for actively traded stocks. Actively traded stocks are often large in market capitalization and are the crucial sample observations in most studies whether they are in investment or corporate finance. Omitting these observations can potentially affect the conclusions of studies. Furthermore, judging from the dominance of *ECIN* over *PIN* in asset pricing tests, *ECIN* is a more precise and preferable measure of information risk that can be utilized in future studies and tests that involve information risk.

Table 1-1 Summary Statistics

This table contains means, medians, and 10th (P10) and 90th (P90) percentiles of values on the variables in the sample of the asset pricing tests. *RET* is the percentage monthly return in excess of the one-month T-bill rate. *ECIN* is information risk measure estimated over the previous year with the model described in Section 3. *PIN* is the probability of informed trading estimated over the previous year with the method described in Easley et al. (2002). *BETA* is the post-ranking portfolio betas of Fama-French (1992) estimated over the previous year using 40 portfolios. *LNMV* is the logarithm of the previous year-end market value of equity (in million dollars). *LNBM* is the logarithm of the previous year-end book value divided by market value. *R6* is the cumulative return of the past six months beginning in the second to last month. *SPR* is the average of daily opening spread over the previous year. *STD* is the standard deviation of daily return in the previous year. *LNTURN* is the natural logarithm of the share turnover measured by the number of shares traded divided by the number of shares outstanding in the second to last month. *LNCVT* is the natural logarithm of the coefficient of variation of turnover calculated over the past 36 months beginning in the second to last month. *ILLIQ* is Amihud (2002) illiquidity measure estimated over the previous year. The sample includes NYSE and Amex common stocks. The sample period is from January 1984 through December 2007 for *RET*, and from January 1983 through December 2006 for all the other variables.

	MEAN	MEDIAN	P10	P90
RET (%)	1.170	0.609	-10.419	12.861
ECIN	-0.246	-0.226	-0.451	-0.068
PIN	0.193	0.176	0.110	0.300
BETA	1.193	1.154	0.800	1.630
LNMV	6.078	6.116	3.334	8.667
LNBM	-0.555	-0.507	-1.426	0.272
R6 (%)	7.052	4.560	-22.464	36.482
SPR (%)	3.214	2.767	1.781	5.265
STD (%)	2.224	2.038	1.220	3.418
LNTURN	-3.060	-2.963	-4.343	-1.940
LNCVT	-0.668	-0.693	-1.209	-0.097
ILLIQ (106)	1.819	0.037	0.001	2.158

Table 1-2 Large and Small Trades**Panel A Dollar-Based Trade-Size Cut-off Points**

This table contains the dollar-based trade-size cut-off points for large and small trades in each size quintile. In each month t , we sort stocks into quintile based on firm size and find the 99th percentile stock price within each quintile. The time-series average of the 99th percentile stock price over the whole sample period is denoted as P99. For each firm size quintile, the cut-off points are $\$100 \times P99$ rounded up to the nearest \$100. The sample includes NYSE and Amex stocks from January 1983 through December 2006 and is divided to two sub-periods.

Sample Period	Small	2	3	4	Large
Year 1983-1994	3,900	4,900	8,300	8,800	17,700
Year 1995-2006	5,100	6,800	10,300	17,000	19,900

Panel B Summary Statistics of Share-Based Trade-Size Cut-off Points

This table summarizes the mean of share-based trade-size cut-off points and average percentages of small, large and block trades for each size quintile. For each stock, we compare its month $t-1$ closing price to the dollar-based cut-off points in panel A and determine the largest number of round lot shares that is less than or equal to that dollar value. This number becomes the share-based cut-off point for the stock over month t . Small trades are trades below the cut-off size, large trades are trades above the cut-off size but below 10,000 shares, and block trades are trades with 10,000 shares and above. The table reports the time-series average of the monthly cross-sectional mean of the share cut-off distribution and the average percentages of small, large and block trades out of the total number of trades. The sample includes NYSE and Amex stocks from January 1983 through December 2006.

Firm-Size Quintile	Avg. Share Cut-off	Percentage of Small Trades	Percentage of Large Trades	Percentage of Block Trades
Year 1983-1994				
Small	400	50.73%	48.68%	0.59%
2	300	45.70%	52.99%	1.31%
3	300	51.83%	46.05%	2.12%
4	300	45.21%	52.13%	2.66%
Large	400	51.47%	45.69%	2.84%
Year 1995-2006				
Small	400	58.50%	40.86%	0.64%
2	300	58.80%	40.15%	1.05%
3	400	59.53%	39.04%	1.43%
4	500	62.95%	35.12%	1.93%
Large	400	52.40%	45.08%	2.52%

Table 1-3 Temporary and Permanent Price Effects of Large and Small Trades

This table reports the means of temporary, permanent, and total price effects of large and small trades. P_s is the logarithm of the trading price. The permanent price effect of trade s is $P_{s+j} - P_{s-i}$; the temporary price effect is $P_s - P_{s+j}$; and the total price effect is $P_s - P_{s-i}$. The t -statistics are reported in parentheses. We reports the cases with $(i=1, j=1)$ and $(i=3, j=3)$ respectively. Panel A reports the effect of seller-initiated trades and Panel B reports the effect of buyer-initiated trades. The sample includes NYSE and Amex stocks from January 1983 through December 2006.

Panel A Seller-Initiated Trades

	Large	Small	Large-Small
<hr/> i=1, j=1 <hr/>			
Permanent Price Effect (%)	-0.613 (-55.91)	-0.551 (-31.86)	-0.062 (-3.50)
Temporary Price Effect (%)	-0.234 (-25.46)	-0.252 (-41.57)	0.018 (1.84)
Total Price Effect (%)	-0.847 (-60.88)	-0.803 (-42.92)	-0.043 (-4.09)
<hr/> i=3, j=3 <hr/>			
Permanent Price Effect (%)	-0.459 (-33.33)	-0.253 (-20.28)	-0.207 (-13.80)
Temporary Price Effect (%)	-0.236 (-23.64)	-0.258 (-29.43)	0.022 (1.89)
Total Price Effect (%)	-0.695 (-58.19)	-0.510 (-52.00)	-0.185 (-18.23)

Panel B Buyer-Initiated Trades

	Large	Small	Large-Small
<hr/>			
i=1, j=1			
<hr/>			
Permanent Price Effect (%)	0.620 (72.68)	0.556 (55.28)	0.064 (7.98)
Temporary Price Effect (%)	0.176 (38.98)	0.255 (8.73)	-0.080 (-2.73)
Total Price Effect (%)	0.796 (75.71)	0.812 (26.08)	-0.016 (-0.54)
<hr/>			
i=3, j=3			
<hr/>			
Permanent Price Effect (%)	0.596 (51.59)	0.340 (34.50)	0.256 (26.15)
Temporary Price Effect (%)	0.107 (18.21)	0.198 (16.01)	-0.091 (-7.31)
Total Price Effect (%)	0.703 (64.04)	0.538 (43.48)	0.164 (13.41)
<hr/>			

Table 1-4 VECM Parameters

This table contains the pooled cross-section time-series summary statistics of the parameter estimates of the following VECM:

$$\Delta P_{i,t}^L = c_i^L + \alpha_i^L (P_{i,t-1}^L - P_{i,t-1}^S) + \sum_{h=1}^q \phi_{i,h} \Delta P_{i,t-h}^L + \sum_{h=1}^q \delta_{i,h} \Delta P_{i,t-h}^S + \eta_{i,t}^L$$

$$\Delta P_{i,t}^S = c_i^S + \alpha_i^S (P_{i,t-1}^L - P_{i,t-1}^S) + \sum_{k=1}^q \lambda_{i,k} \Delta P_{i,t-k}^L + \sum_{k=1}^q \gamma_{i,k} \Delta P_{i,t-k}^S + \eta_{i,t}^S$$

where P^L is the price of the large trade and P^S is the price of the small trade. The model is estimated for each firm-year on NYSE and Amex stocks from 1983 through 2006. We choose $q=5$ for the autoregressive lags. Panel A reports the estimated parameters from VECM. α^L measures the price discovery of the large trades, and α^S measures the price discovery of the small trades. σ^L is the standard deviation of the error term η^L , and σ^S is the standard deviation of the error term η^S . Panel B reports the characteristics of large and small prices. The mean (AVG) and standard deviation (STD) of $(P_i^L - P_i^S)$, ΔP_i^L , and ΔP_i^S are estimated for each firm-year. Then the pooled cross-section time-series summary statistics are reported in percentage. $P10$, $P25$, $P75$ and $P90$ are 10th, 25th, 75th and 90th percentiles, respectively. Panel C gives summary statistics of *ECIN* by year.

Panel A VECM Parameters

	MEDIAN	P10	P25	P75	P90
α^L	-0.172	-0.513	-0.316	-0.095	-0.060
α^S	0.182	0.052	0.091	0.359	0.516
σ^L	0.123	0.067	0.088	0.180	0.285
σ^S	0.121	0.065	0.086	0.178	0.284

Panel B Characteristics of Large and Small Prices

	MEDIAN	P10	P25	P75	P90	Correlation with α^L
AVG $(P_i^L - P_i^S)$	-0.046	-1.426	-0.585	0.302	1.048	-0.087
AVG (ΔP_i^L)	0.022	-0.300	-0.111	0.159	0.355	-0.013
AVG (ΔP_i^S)	0.000	-1.069	-0.418	0.366	0.810	-0.257
STD $(P_i^L - P_i^S)$	0.104	0.034	0.071	0.144	0.197	0.230
STD (ΔP_i^L)	0.125	0.068	0.089	0.187	0.307	-0.167
STD (ΔP_i^S)	0.293	0.116	0.174	0.508	0.850	0.050

Panel C Summary Statistics of *ECIN* by Year

Year	N	MEAN	STD	P10	P25	MEDIAN	P75	P90
1983	1695	-0.138	0.095	-0.248	-0.173	-0.118	-0.082	-0.056
1984	1669	-0.119	0.100	-0.221	-0.148	-0.101	-0.066	-0.036
1985	1602	-0.142	0.107	-0.268	-0.186	-0.114	-0.079	-0.052
1986	1594	-0.166	0.125	-0.305	-0.227	-0.134	-0.088	-0.060
1987	1576	-0.167	0.131	-0.307	-0.225	-0.140	-0.090	-0.053
1988	1236	-0.150	0.104	-0.300	-0.214	-0.119	-0.078	-0.051
1989	1424	-0.163	0.122	-0.316	-0.232	-0.135	-0.081	-0.053
1990	1381	-0.153	0.125	-0.299	-0.222	-0.124	-0.075	-0.048
1991	1246	-0.176	0.122	-0.318	-0.249	-0.152	-0.088	-0.059
1992	1385	-0.178	0.120	-0.322	-0.248	-0.148	-0.094	-0.065
1993	1203	-0.174	0.115	-0.324	-0.243	-0.143	-0.093	-0.064
1994	1302	-0.175	0.111	-0.332	-0.252	-0.146	-0.094	-0.066
1995	1389	-0.174	0.115	-0.327	-0.244	-0.151	-0.090	-0.063
1996	1452	-0.183	0.117	-0.337	-0.260	-0.160	-0.095	-0.063
1997	1610	-0.203	0.132	-0.374	-0.280	-0.179	-0.105	-0.068
1998	1760	-0.217	0.259	-0.419	-0.310	-0.193	-0.103	-0.061
1999	1643	-0.240	0.172	-0.463	-0.343	-0.209	-0.108	-0.065
2000	1561	-0.274	0.202	-0.532	-0.381	-0.238	-0.114	-0.070
2001	1331	-0.330	0.228	-0.641	-0.489	-0.309	-0.132	-0.065
2002	1371	-0.381	0.254	-0.708	-0.543	-0.383	-0.163	-0.077
2003	1389	-0.431	0.274	-0.797	-0.620	-0.417	-0.211	-0.086
2004	1464	-0.428	0.281	-0.780	-0.596	-0.415	-0.225	-0.091
2005	1473	-0.459	0.305	-0.833	-0.643	-0.442	-0.230	-0.092
2006	1470	-0.463	0.299	-0.829	-0.640	-0.453	-0.259	-0.091
Avg.	1468	-0.237	0.167	-0.442	-0.332	-0.214	-0.118	-0.065

Table 1-5 Characteristics of *ECIN*

Panel A Simple Correlations

This table reports the time-series means of the cross-sectional correlations of the variables. RET is the percentage monthly return in excess of the one-month T-bill rate. $ECIN$ is information risk measure estimated over the previous year with the model described in Section 3. PIN is the probability of informed trading estimated over the previous year with the method described in Easley et al. (2002). $BETA$ is the post-ranking portfolio betas of Fama-French (1992) estimated over the previous year using 40 portfolios. $LNMV$ is the logarithm of the previous year-end market value of equity (in million dollars). $LNBM$ is the logarithm of the previous year-end book value divided by market value. $R6$ is the cumulative return of the past six months beginning in the second to last month. SPR is the average of daily opening spread over the previous year. STD is the standard deviation of daily return in the previous year. $LNTURN$ is the natural logarithm of the share turnover measured by the number of shares traded divided by the number of shares outstanding in the second to last month. $LNCVT$ is the natural logarithm of the coefficient of variation of turnover calculated over the past 36 months beginning in the second to last month. $ILLIQ$ is Amihud (2002) illiquidity measure estimated over the previous year. The sample includes NYSE and Amex common stocks. The sample period is from January 1984 through December 2007 for RET , and from January 1983 through December 2006 for all the other variables.

[illegible]

Panel B Quintile Analysis

The table reports the yearly average of the cross-section means of *ECIN*, *LNMV*, *SPR*, and *PIN* in each *ECIN* quintiles, together with the minimum and maximum of *ECIN* within each quintile. *ECIN* is information risk measure estimated yearly for each stock. *LNMV* is the logarithm of the year-end market value of equity (in million dollars). *SPR* is the average of daily opening spread over the year. *PIN* is the probability of informed trading estimated with the method described in Easley et al. (2002). The sample includes NYSE and Amex common stocks from January 1983 through December 2006.

<i>ECIN</i> Quintile	Mean <i>ECIN</i>	Min. <i>ECIN</i>	Max. <i>ECIN</i>	<i>LNMV</i>	<i>SPR</i> (%)	<i>PIN</i>
Low	-0.479	-1.196	-0.368	8.112	2.796	0.139
2	-0.312	-0.367	-0.262	7.175**	2.986**	0.159**
3	-0.220	-0.262	-0.179	6.273**	3.176**	0.183**
4	-0.142	-0.178	-0.104	5.467**	3.469**	0.209**
High	-0.055	-0.104	0.801	4.690**	3.802**	0.250**

** indicates the difference from the previous quintile is statistically significant at 1% level.

Table 1-6 Asset Pricing Tests

This table reports the time-series averages of coefficients in the cross-sectional asset-pricing tests using standard Fama and MacBeth (FM; 1973) methodology and Litzenberger and Ramaswamy (1979) weighted least-squares (LR-WLS) procedure. The cross-sectional regression is estimated for each month in the sample period over year 1984 to 2007. Then the coefficients of cross-sectional regressions are averaged through time and *t*-statistics (in the parentheses) of the time-series means are calculated accordingly. The dependent variable is the month *t* return (in percentage) in excess of one-month T-bill rate. *ECIN* is the information risk measure estimated in the previous calendar year. *BETA* is the post-ranking portfolio betas of Fama-French (1992) in the previous year estimated based on 40 portfolios. *LNBMV* is the logarithm of the market value of equity at the previous year end (in million dollars). *LNBM* is the logarithm of the book value divided by the market value at the previous year end. *R6* is the six-month cumulative return from month *t*-2 to *t*-7. *ILLIQ* is the Amihud (2002) illiquidity measure in 10^6 estimated in the previous year.

		<i>ECIN</i>	<i>BETA</i>	<i>LNBM</i>	<i>LNBMV</i>	<i>ILLIQ</i>	<i>R6</i>
(1)	FM		-0.095 (-0.33)	0.161 (1.72)	-0.012 (-0.28)		
	LR-WLS		-0.603 (-2.60)	0.198 (2.45)	0.028 (0.78)		
(2)	FM	0.919 (3.47)	-0.078 (-0.27)	0.128 (1.42)	0.120 (3.11)		
	LR-WLS	0.444 (3.30)	-0.425 (-1.70)	0.151 (1.86)	0.112 (3.42)		
(3)	FM	0.753 (3.07)	-0.132 (-0.47)	0.169 (1.95)	0.097 (2.55)		1.170 (4.39)
	LR-WLS	0.322 (2.30)	-0.447 (-1.82)	0.194 (2.44)	0.091 (2.86)		1.325 (6.07)
(4)	FM		-0.090 (-0.32)	0.163 (1.74)	0.031 (0.75)	0.033 (3.13)	
	LR-WLS		-0.591 (-2.56)	0.200 (2.47)	0.056 (1.61)	0.014 (3.87)	
(5)	FM	0.894 (3.33)	-0.081 (-0.28)	0.128 (1.40)	0.122 (2.97)	0.056 (0.81)	
	LR-WLS	0.422 (3.10)	-0.418 (-1.69)	0.150 (1.85)	0.117 (3.32)	0.014 (0.28)	

Table 1-7 Asset Pricing Tests—Further Checks

This table reports the results from the monthly Fama-MacBeth regressions. The cross-sectional regression is estimated for each month in the sample period over year 1984 to 2007. Then the coefficients of cross-sectional regressions are averaged through time and t-statistics (in the parentheses) of the time-series means are calculated accordingly. The dependent variable is the month t return (in percentage) in excess of the one-month T-bill rate. *ECIN* is the information risk measure estimated in the previous calendar year. *BETA* is the post-ranking portfolio betas of Fama-French (1992) in the previous year estimated based on 40 portfolios. *LNBMV* is the logarithm of the market value of equity at the previous year end (in million dollars). *LNBM* is the logarithm of the book value divided by the market value at the previous year end. *SPR* is the average of daily opening spread over the previous year. *STD* is the standard deviation of daily return in the previous year. *LNTURN* is the natural logarithm of the share turnover measured by the number of shares traded divided by the number of shares outstanding in the second to last month. *LNCVT* is the natural logarithm of the coefficient of variation of turnover calculated over the past 36 months beginning in the second to last month. *No_LS* is the percentage of time-intervals without large-small trade pairs in the total number of observations fitted in VECM. *PECIN* is an instrument variable created to test the EIV problem. Regression (1), (3), (5), (7) and (9) are run in non-January months, while Regression (2), (4), (6), (8) and (10) are run in Januarys.

	<i>ECIN</i>	<i>BETA</i>	<i>LNBM</i>	<i>LNBMV</i>	<i>LNTURN</i>	<i>LNCVT</i>	<i>SPR</i>	<i>STD</i>	<i>R6</i>	<i>ILLIQ</i>	<i>PECIN</i>	<i>NO_LS</i>
(1)	1.337 (4.59)	-0.328 (-1.09)	0.03 (0.34)	0.175 (4.31)								
(2)	-3.922 (-3.62)	2.79 (3.20)	1.132 (3.23)	-0.503 (-3.46)								
(3)	0.911 (3.82)	-0.362 (-1.33)	0.064 (0.75)	0.101 (2.46)	-0.036 (-0.47)	-0.265 (-2.53)						
(4)	-3.635 (-2.65)	2.511 (3.14)	1.146 (3.37)	-0.501 (-3.72)	0.293 (1.33)	0.257 (0.82)						
(5)	0.940 (3.28)	-0.357 (-1.20)	0.087 (1.00)	0.124 (3.04)			-3.782 (-0.73)					
(6)	-3.308 (-3.45)	2.472 (3.04)	1.078 (3.21)	-0.181 (-1.22)			79.642 (3.30)					
(7)	0.473 (2.27)	0.090 (0.42)	-0.035 (-0.44)	0.064 (1.64)				-0.439 (-4.48)				
(8)	-1.980 (-1.63)	1.444 (2.73)	1.289 (3.63)	-0.236 (-2.45)				1.032 (3.35)				
(9)	0.346 (1.69)	-0.152 (-0.78)	0.033 (0.46)	0.062 (1.57)	-0.024 (-0.43)	-0.217 (-2.42)	4.459 (0.91)	-0.272 (-2.91)	1.148 (4.07)	0.110 (1.52)		
(10)	-2.094 (-1.65)	1.360 (2.63)	1.191 (4.08)	-0.056 (-0.45)	0.023 (0.13)	-0.164 (-0.85)	45.51 (2.34)	0.863 (2.94)	0.682 (0.77)	0.553 (2.62)		
(11)	0.441 (1.81)	0.045 (0.15)	0.143 (1.49)	0.090 (2.18)							0.251 (0.52)	
(12)	0.701 (2.57)	-0.102 (-0.34)	0.162 (1.80)	0.090 (2.16)								-0.096 (-0.36)

Table 1-8 Asset Pricing Tests – Comparing *ECIN* with *PIN*

This table reports the results from the monthly Fama-MacBeth regressions. The cross-sectional regression is estimated for each month in the sample period over year 1984 to 2007. Then the coefficients of cross-sectional regressions are averaged through time and *t*-statistics (in the parentheses) of the time-series means are calculated accordingly. The dependent variable is the month *t* return (in percentage) in excess of the one-month T-bill rate. *ECIN* is the information risk measure estimated in the previous calendar year. *PIN* is the probability of informed trading estimated over the previous year with the method described in Easley et al. (2002). *BETA* is the post-ranking portfolio betas of Fama-French (1992) in the previous year estimated based on 40 portfolios. *LNBMV* is the logarithm of the market value of equity at the previous year end. *LNBM* is the logarithm of the book value divided by the market value at the previous year end. *ILLIQ* is the Amihud (2002) illiquidity measure in 10^6 estimated in the previous year.

	<i>ECIN</i>	<i>PIN</i>	<i>BETA</i>	<i>LNBM</i>	<i>LNBMV</i>	<i>ILLIQ</i>
(1)		1.103 (1.67)	-0.115 (-0.40)	0.127 (1.45)	0.111 (2.78)	
(2)	0.811 (3.10)	0.975 (1.51)	-0.089 (-0.32)	0.136 (1.56)	0.144 (3.25)	
(3)		0.487 (0.80)	-0.099 (-0.36)	0.152 (1.65)	0.042 (0.90)	0.034 (3.19)
(4)	0.822 (3.04)	0.965 (1.53)	-0.068 (-0.24)	0.127 (1.45)	0.146 (3.26)	0.044 (0.63)

Table 1-9 Economic Significance of *ECIN*

Panel A Raw Return

This table reports time-series averages of coefficients in the cross-sectional asset-pricing tests using standard Fama and MacBeth (FM; 1973) methodology and Litzenberger and Ramaswamy (1979) weighted least-squares (LR-WLS). The cross-sectional regression is estimated for each month in the sample period over year 1984 to 2007. Then the coefficients of cross-sectional regressions are averaged through time and *t*-statistics (in the parentheses) of the time-series means are calculated accordingly. The dependent variable is the month *t* return (in percentage). *HECIN* (*LECIN*) is the high (low) information risk dummy variable that takes the value of one if the stock ranks among the top (bottom) quintile in *ECIN* estimated over the previous year and take the value of zero otherwise. *BETA* is the post-ranking portfolio betas of Fama-French (1992) in the previous year estimated based on 40 portfolios. *LNMV* is the logarithm of market value of equity at the previous year end. *LNBM* is the logarithm of the book value divided by the market value at the previous year end. *R6* is the six-month cumulative return from month *t*-2 to *t*-7. *BETA*, *LNMV*, *LNBM* and *R6* have been demeaned in this test.

	FM		LR-WLS	
	Coefficients	<i>t</i> -statistic	Coefficients	<i>t</i> -statistic
<i>INTERCEPT</i>	1.130	3.94	0.904	3.72
<i>HECIN</i>	0.030	0.33	0.127	1.60
<i>LECIN</i>	-0.142	-2.42	-0.138	-2.57
<i>BETA</i>	-0.149	-0.54	-0.458	-2.13
<i>LNMV</i>	0.084	2.21	0.084	2.39
<i>LNBM</i>	0.162	1.87	0.187	2.73
<i>R6</i>	1.178	4.40	1.336	5.89

Panel B Risk-Adjusted Return

This table reports results of the time-series regressions where we regress HL , which is the time series difference between the coefficients on $HECIN$ and $LECIN$ from the FM regression or LR-WLS regression conducted in Panel A, on the contemporaneous Fama-French four factors and liquidity factors. MKT is the market excess return, SMB is the small-minus-big factor, HML is the high-minus-low factor, and UMD is the momentum factor. $PS-VWF$ is the value-weighted Pastor-Stambaugh Liquidity traded factor. $SDK-VWF$ is the value-weighted Sadka liquidity traded factor. The sample period of this test is from January 1984 through December 2007.

	HL_{FM}			HL_{LR-WLS}		
	(1)	(2)	(3)	(4)	(5)	(6)
<i>INTERCEPT</i>	0.290 (2.63)	0.294 (2.66)	0.288 (2.54)	0.385 (4.25)	0.378 (4.17)	0.369 (3.91)
<i>MKT</i>	-19.964 (-7.35)	-20.234 (-7.33)	-20.993 (-7.28)	-15.533 (-6.96)	-15.035 (-6.64)	-15.929 (-6.63)
<i>SMB</i>	-16.347 (-4.87)	-16.173 (-4.79)	-18.163 (-5.23)	-10.762 (-3.90)	-11.084 (-4.00)	-13.536 (-4.69)
<i>HML</i>	-7.437 (-1.84)	-7.917 (-1.91)	-7.919 (-1.90)	-6.130 (-1.84)	-5.246 (-1.55)	-7.251 (-2.09)
<i>UMD</i>	9.769 (4.13)	10.444 (3.95)	9.283 (3.72)	5.326 (2.74)	4.080 (1.88)	4.183 (2.01)
<i>PS-VWF</i>		-1.270 (-0.57)			2.342 (1.29)	
<i>SDK-VWF</i>			0.651 (0.23)			1.687 (0.71)

Figure 1-1 Cumulative Average Returns in Seller (Buyer)-Initiated Trades

This figure is constructed in the following manner. A large (small) trade with a seller (buyer)-initiated trade indicator is designed as trade 0. The previous 10 trades (regardless trade size and trade direction) prior to trade 0 are trades -1, -2, and so on. The trades after trade 0 are 1, 2, and so on. Trade-to-trade log-returns are cumulated from trade -10 to trade +10. Then the returns are averaged across the whole sample.

2. INFORMATION RISK FACTOR AND ASSET PRICING ANOMALIES

2.1 Introduction

Traditional asset pricing theory assumes that information risk is potentially idiosyncratic and completely diversifiable. However, a series of papers by Easley and O'Hara (e.g. Easley, Hvidkjaer and O'Hara (2002), Easley and O'Hara (2004), and O'Hara (2003)) suggests that information risk is systematic and undiversifiable. Following Grossman and Stiglitz (1980), who argue that an informational fully-revealing equilibrium is impossible in a real market, Easley and O'Hara (2004) develop the theoretical work that the uninformed trader has to hold an undiversifiable information risk when she trades with the privately informed trader. The uninformed trader can only avoid information risk by holding nothing but the risk-free asset, which is nevertheless not optimal for the rational investor who aims to maximize her utility. Furthermore, without the informed investor's private information, the uninformed investor cannot form portfolios to fully diversify information risk, so she rationally requires compensation for bearing such risk. Easley, Hvidkjaer and O'Hara (2002) provide empirical evidence that information risk proxy PIN is cross-sectionally priced as a firm characteristic. In addition, Easley, Hvidkjaer and O'Hara (2005) find that the zero-investment portfolio, which is size-neutral but takes a long position on high- PIN stocks and a short position on low- PIN stocks, earns a significant abnormal return that is adjusted for the Fama-French and momentum risk factors. Based on this observation, they suggest that PIN is a proxy for some underlying information risk

factor. However, Hughes, Liu and Liu (2007) and Lambert, Leuz and Verrecchia (2007) argue that information risk is either fully diversifiable when the economy is large enough or subsumed by existing risk factors. In addition, Mohanram and Rajgopal (2008) find that the pricing effect of *PIN* is not robust to alternative time periods and that the loadings on the *PIN* factor are not priced in two-stage cross-sectional asset pricing tests. Duarte and Young (2009) find *PIN* to be subsumed by the illiquidity measure in the asset pricing test.

Due to the absence of an adequate information risk measure, we still lack compelling evidence on information risk's being priced. The first essay creates a new information risk measure *ECIN* based on the price discovery of large trades and illustrates that *ECIN* is priced significantly in cross-sectional returns. The pricing effect of *ECIN* subsumes *PIN*, and is also robust in different sample periods and to various illiquidity measures. In this essay, we create an information risk factor based on this *ECIN* measure and provide evidence that the *ECIN* factor is a priced common risk factor. Similar to Easley, Hvidkjaer and O'Hara (2005), we form 10 zero-investment portfolios on firm size deciles, which take long positions on high-*ECIN* stocks and short positions on low-*ECIN* stocks. The monthly average return of these size-neutral *ECIN* portfolios is both economically and statistically significant and cannot be explained by the three Fama and French (1993) risk factors or the Carhart (1997) momentum factor. This unexplained abnormal return together with the strong covariation between the *ECIN* portfolios suggests that *ECIN* might be a proxy for an underlying risk factor.

We create the *ECIN* factor-mimicking portfolio as the monthly average return of 10 zero-investment size-neutral portfolios and denote it as *ECINF*. We explicitly test whether *ECINF* is a priced risk factor using a two-stage cross-sectional regression approach (2SCSR) that estimates the factor loadings on the first stage and the factor risk premiums on the second stage. Under this method, the significance of the factor risk premium uncovers whether the risk factor is priced. This method has been widely used to test whether a candidate variable is a priced risk factor, for example, to test the CAPM (Fama and MacBeth (1973)), the conditional CAPM (Jagannathan and Wang (1996)), the inter-temporal CAPM (Brennan, Wang and Xiao (2004), Petkova (2006)), and the two-beta model (Campbell and Vuolteenaho (2004)). This method is also used to test whether the accrual is a priced risk factor (Core, Guay and Verdi (2008), Hirshleifer, Hou and Teoh (2006), Khan (2008)) and whether *PIN* is a priced risk factor (Mohanram and Rajgopal (2008)). In this paper, we run 2SCSR tests to examine whether *ECINF* is a priced risk factor. The tests are conducted on three different sets of portfolios: the first set consists of 25 portfolios sorted independently by firm size and *ECIN*; the second set consists of 27 portfolios sorted independently by firm size, book-to-market ratio and *ECIN*; the third set consists of the 25 Fama-French portfolios sorted independently by firm size and book-to-market ratio. The 25 Fama-French portfolios are constructed on NYSE/Amex/Nasdaq stocks, while *ECINF* is constructed based only on NYSE/Amex stocks. The 2SCSR test on this set of portfolio gives an out-of-sample test of whether *ECINF* is a commonly priced risk factor. The 2SCSR tests on all three sets of portfolios reveal significantly positive risk premium of *ECINF*, indicating that *ECINF* is a priced risk factor. It is interesting to contrast our results of 2SCSR tests with those in Mohanram and Rajgopal (2008) where *PIN* factor loadings are insignificantly related to the cross-sectional returns.

This sharp contrast suggests that *ECINF* is superior to the *PIN* factor in capturing underlying information risk.

As the 2SCSR tests suggest that *ECINF* is a priced risk factor, we further exploit it to explain the financial anomalies which are potentially related to information asymmetry. Financial anomalies are price patterns violating the efficient market assumption, under which any lags in the stock prices' response to an event are short-lived. The growing literature on such financial anomalies reveals that stock prices adjust slowly to information by drifting over time in the same direction as their initial surprise. The long-term underreaction or overreaction to information results in systematic patterns in long-term stock returns following public news. The traditional asset pricing models have a hard time explaining these anomalies, while behavioral alternatives (e.g. Barberis, Shleifer and Vishny (1998) and Daniel, Hirshleifer and Subrahmanyam (1998)) have some success. However, behavioral models rely on several psychology theories, and Fama (1998) criticizes that although the mixed and matched behavioral models can explain some specific anomalies, we lack one efficient model to capture the whole picture of all anomalies.

Hong and Stein (1999) explain financial anomalies based on gradual information flow. They develop their model on the assumption of heterogeneous agents who process only subset of information. Private information diffuses gradually across the population of investors, leading to an initial under-reaction and a following over-reaction. In addition, the Hong and Stein (1999) model indicates that momentum strategies should be most profitable among stocks in which information moves most slowly across the investing public. Brav and Heaton (2002) provide the theoretical

explanation to the financial anomalies assuming that investors are rational and have incomplete information sets. Caskey (2008) attributes the financial anomalies to failure to fully impound public information into stock prices. These information-based explanations motivate us to use information risk factor constructed in this paper to explain financial anomalies. As truly and purely public information should cause stock prices to change immediately without any trade, the existing long-term trend in stock prices raises the possibility that public information events may have some private component due to different abilities of traders in interpreting or utilizing public information. Sadka (2006) suggests that momentum and post-earnings-announcement drift (PEAD) returns can be viewed as compensation for “unexpected variations in the aggregate ratio of informed traders to noise traders”.

Fama (1998) argues that most long-term return anomalies tend to disappear with reasonable changes in methodology, but only the price and earnings momentum are robust and persistent open puzzles. This observation prompts us to focus our attention on price momentum, and post-earnings announcement drift. We also study the stock issuance puzzle that is newly documented by Fama and French (2008) and Pontiff and Woodgate (2008), and the closely related anomaly that describes the abnormal long-run underperformance in stock returns following seasoned equity offerings (SEO), commonly referred to as SEO underperformance. In the following paragraphs, we describe the above-mentioned four anomalies in turn and provide intuitions for how each anomaly can be related to information risk.

Price momentum, documented by Jegadeesh and Titman (1993), refers to the phenomenon in which one can obtain superior returns by holding a zero-investment portfolio that takes a long position in past winners and a short position in past losers.

The abnormal return generated by this momentum strategy cannot be explained by the traditional risk factors (the market, size, and value factors of Fama and French (1993)). Cohen, Gompers and Vuolteenaho (2002) show that stocks with low institutional ownership exhibit high momentum. As higher institutional ownership is typically associated with higher information dissemination, this evidence suggests that information diffusion or dissemination plays a role in momentum. Hong and Stein (1999) attribute the momentum anomaly to the slow diffusion of private information. In their model there are two groups of investors, which are news-watchers and momentum traders. Private information diffuses gradually across the news-watchers, generating the return trend. The momentum traders chase the trend and create additional return continuation. As observed by Hong, Lim and Stein (2000), stocks with slower information diffusion exhibit more pronounced momentum and this is especially manifest in the past losers rather than in the past winners. Slower diffusion of information in the losers suggests that there is less informed trading and hence smaller information risk in the losers, which can potentially explain their lower returns. Short-sale constraints also make informed arbitrage especially difficult in bad news firms (Diamond and Verrecchia (1987), Shleifer and Vishny (1997), Hong, Lim and Stein (2000) and Chan (2003).) Therefore, losers earn lower return and can be considered as a hedge against information asymmetry. Sadka (2006) finds that winners (losers) have positive (negative) loadings on information-related liquidity risk factors. We find a similar monotonic decreasing pattern of loadings on *ECINF* from positive on the winners to negative on the losers. In addition, our empirical results show that the returns of zero-investment momentum portfolios, which are long on

winners and short on losers, have significantly positive loadings on *ECINF*, and are insignificantly different from zero after adjusting for information risk.

Earnings momentum or post-earnings-announcement drift (PEAD) refers to the phenomenon that a zero-investment portfolio with a long position on the high standardized-unexpected earning (SUE) firms and a short position on the low SUE firms generates superior risk-adjusted returns for up to 12 months after the portfolio formation. This anomaly has been widely studied since Ball and Brown (1968). Similar to the price momentum, the post-earnings-announcement drift is also related to information imperfections (e.g. Francis, Lafond, Olsson and Schipper (2007) and Vega (2006)). Bernard and Thomas (1990) suggest that investors form their earnings expectations according to the seasonal random walk (SRW) model, which has prediction errors exhibiting a serial correlation. The ignorance of this autocorrelations leads to the PEAD. Soffer and Lys (1999) find that information dissemination can revise investors' expectation to the level that implicitly incorporates the serial correlation of earning surprise. Wang (2008) documents that better accounting disclosure can mitigate PEAD by helping investors incorporate both prior and future earnings information. Ng, Rusticus and Verdi (2007) argue that transaction costs hinder informed trades from incorporating earnings information into price, and that there is a positive relation between the transaction cost and PEAD. More recently, Livnat and Mendenhall (2006) document that analyst forecast surprises may capture some different form of mispricing than the SRW model. Battalio and Mendenhall (2005) find that investors of different trading sizes react differently to analyst forecast errors and to SRW model errors. We use both the SRW model and the analyst forecast to estimate SUE, and find that information risk can explain the PEAD measured by both of these two methods. As the same factors that contribute to the lower

information risk for loser stocks in price momentum would also make information risk lower in the low SUE stocks relative to the high SUE stocks, the lower information risk premium can potentially explain the lower returns of the low SUE stocks and PEAD.

We find that the zero-investment SUE portfolios have significantly positive loading on information risk factor *ECINF* and earn insignificant information risk adjusted returns. This result suggests that the PEAD, similar to the price momentum, reflects the different exposures to information risk caused by asymmetric informed trading between good news and bad news firms. A closer look at the factor loadings on each SUE quintile suggests that low (high) SUE firms have negative (positive) loadings on *ECINF*. The negative loadings on *ECINF* suggest that bad news firms are negatively exposed (a hedge) to information risk, due to the fact that short-sale costs and restrictions constrain informed trading on the selling side when there is bad news. Similarly, Sadka (2006) finds increasing loadings from the lowest SUE firms to the highest SUE firms on the information-risk-related illiquidity factor.

The underperformance of SEO firms was documented by Loughran and Ritter (1995) and Spiess and Affleck-Graves (1995). Several theories attempt to explain this new issue puzzle. Teoh, Welch and Wong (1998) and Rangan (1998) address the relation between SEO long-run underperformance and earnings management. Issuers manipulate current accruals to report higher net income prior to the offering and subsequently disappoint the market with predictable declines in earnings. However, Shivakumar (2000) argues that investors can fully anticipate the earnings management prior to SEO and rationally undo this effect at equity offering announcement. Daniel,

Hirshleifer and Subrahmanyam (1998) provide the behavioral explanation that the post-event price drift is caused by investors' overconfidence. Investors overreact to recent trends in stock performance when managers issue equity in order to take advantage of overvalued stock prices, and investors are slow to recognize information conveyed by issue announcements. On the other hand, some argue that the SEO firms underperform because they are less risky. Eckbo, Masulis and Norli (2000) illustrate that equity issuers have lower leverage so their exposure to unexpected inflation and default risks also are lower. In addition, the SEO firms increase stock liquidity and therefore reduce illiquidity risk premium in the post-offering period. Carlson, Fisher and Giammarino (2006) argue that the SEO firms are less risky because the stock issuance are associated with firm investments which convert growth options into assets in place and therefore cause the fundamental change in total asset risk.

We hypothesize that the SEO underperformance is related to information risk for the following reasons. Almazan, Suarez and Titman (2003) argue that raising capital forces firms to disclose some of the internal information, which provides investors a clearer assessment of the firm value. And Lang and Lundholm (2000) find that the SEO firms tend to increase voluntary disclosure to "hype" the stock. In the mean time, raising capital gives outsiders incentive to scrutinize the firm. As such, we can expect that SEO increases public information, decreases the fraction of privately informed traders, and therefore decreases adverse selection costs and information risk. Spiess and Affleck-Graves (1995) suggest that long-run underperformance following equity offering might be associated with the reduced risk in the issuing companies. This argument together with the evidence that information risk is a priced systematic risk suggests that the long-term underperformance in returns can attribute to reduced exposure to information risk in post-SEO years. In order to test this effect, we run the

regression of the returns of SEO firms in pre-SEO and post-SEO years separately on the Fama-French factors and *ECINF*. The returns in pre-SEO years do not have significant loading on *ECINF* while the returns in post-SEO years have significant negative loading on *ECINF*. The abnormal returns in the post-SEO period adjusted to the Fama-French risk factors are fully explained by *ECINF*. Moreover, the yearly average *ECIN* characteristic is significantly smaller in the post-SEO years than that in the pre-SEO year. Therefore, the long-run underperformance of SEO firms may reflect the reduced information risk.

Recently documented by Pontiff and Woodgate (2008) and Fama and French (2008), the stocks with increasing in shares outstanding underperform the stocks with decreasing in shares outstanding. Similar to SEO effects, this share issuance anomaly could also be related to information asymmetry. We create the zero-investment portfolio that takes a long position on the positive issue firms and a short position on the negative issue firms. The significant negative loading on *ECINF* supports the argument that the issuance effect is related to the asymmetric exposure to information risk. In addition, *ECINF* reduces the negative abnormal return to an insignificant level. We then delete the SEO firms from the sample to test whether this result is driven by SEO effects previously studied. The non-SEO positive issue firms still significantly underperformed compared to the negative issue firms; this underperformance can also be captured by the negative loading on the information risk factor. So we conclude that the share issuance effect observed by Pontiff and Woodgate (2008) and Fama and French (2008) can be explained by information risk.

As financial anomalies are news-related, we expect that associated returns are related to shocks of the market-wide information asymmetry. Ignoring information risk raises those unexplainable abnormal returns. The traditional asset pricing model with symmetric information assumptions cannot explain the abnormal return of price and earnings momentum or share issuance effects, while our results suggest that abnormal returns of these anomalies can be captured by information risk. As information risk is undiversifiable systematic risk, it does not suggest market inefficiency as there is no arbitrage opportunity and each portfolio investigated here is well compensated by risk premium. On the contrary, this paper illustrates that failure to incorporate information risk in asset pricing models can produce misleading anomalies which are typically interpreted as evidences against market efficiency.

This paper is organized as follows. Section 2 introduces the data and the estimation of *ECIN*. Section 3 constructs the *ECIN* factor and conducts the two-stage cross-sectional regression tests. Section 4 explores the informational explanation of financial anomalies. Section 5 concludes.

2.2 Data and Estimation of ECIN

The *ECIN* portfolios and factor are formed with all ordinary common stocks listed in New York Stock Exchange (NYSE) and American Stock Exchange (Amex). We obtain the intraday quote and trade data from Institute for the Study of Security Markets (ISSM) and Trade and Quote (TAQ) databases, and the stock returns data from CRSP monthly file. We exclude stocks with less than six months of records in CRSP and stocks with negative or zero shares outstanding. We also delete REIT,

stocks of companies incorporated outside U.S. and closed-end funds. We exclude ISSM/TAQ intraday data with non-positive price, for trades outside regular trading hours, or with irregular terms.⁷ *ECIN* is estimated with ISSM/TAQ data from January 1983 through December 2006; and the *ECIN* factor is formed from January 1984 through December 2007.

The price anomalies analysis is based on common stocks in NYSE, Amex, and Nasdaq. The price momentum and SUE portfolios are formed on NYSE, Amex, and Nasdaq common stocks with valid monthly return data in CRSP. The SUE portfolios also require valid earning announcement dates and earnings per share data in Compustat. The SUE portfolios based on the analysts' forecast method require valid I/B/E/S data. The SEO data is from the Securities Data Corporation (SDC). We include seasoned offerings of common stocks in NYSE, Amex and Nasdaq, which are over 10 million US dollars and over 5% of shares outstanding before the issuing. In the stock issuance analysis, we need shares outstanding and share adjustment factors in CRSP. To eliminate the effect of penny stocks, we also delete stocks if their prices are less than \$5 at the portfolios formation periods.

⁷ We delete trades and quotes outside the regular trading time 9:30-16:00. For ISSM data, we only include trades and quotes with their condition code as blank or “*”. For TAQ quote data, we exclude quotes with their mode as 4,7,9,11,13,14,15,19,20,27,or 28. For TAQ trade data, we only include trades with the correction indicator as 0 or 1 and the condition code as blank or “*”.

As information-based trades are not directly observable, we use the vector error-correction model adopted in the first essay (equation (1.1)) to estimate the *ECIN* for each firm-year as the proxy for information asymmetry risk.

2.3 ECIN Portfolios and ECIN Factor

In this section, we first test whether portfolios sorted by *ECIN* earn diverse returns. We first sort the portfolio with a method same as Easley, Hvidkjaer and O'Hara (2005). At the beginning of each year, we sort all stocks in NYSE and Amex with non-missing *ECIN* data into size deciles according to their previous year-end market capitalizations. Within each size decile, three portfolios are sorted based on their *ECIN* estimated over the previous year. We denote the highest, medium and lowest *ECIN* terciles as *HECIN*, *MECIN* and *LECIN*, respectively. The first three columns in Panel A of Table 2-1 report the average *ECIN* for the stocks in each portfolio and show that the average *ECIN* within each *ECIN* classification decreases monotonically with firm size. The first three columns in Panel B of Table 2-1 report the average size of stocks in each portfolio and show that average size declines monotonically with *ECIN* within each size classification, except for those within the largest size decile. The negative correlation between firm size and *ECIN* that we observe in Panel A and Panel B of Table 2-1 is intuitive since large firms have more informational intermediary coverage and therefore smaller information asymmetry than small firms have. However, this negative correlation also indicates that the dependent sort by *ECIN* is almost a second (reversed) sort by size. To correct this problem, we adopt a method similar to Mohanram and Rajgopal (2008) and Penman (1983): we sort three portfolios based on firm size within each size decile, where this second sort by size is independent of the

sort by *ECIN* in the first step. We denote the largest, medium and smallest size terciles in the second sort as *LSIZE*, *MSIZE* and *SSIZE*, respectively. The characteristics of these portfolios are reported in the same panels with the first *ECIN*-based sorting in the Table 2-1. The *ECIN* portfolio net of size effect is defined as (High *ECIN*- Small Size)-(Low *ECIN*- Large Size)⁸ for each size decile, and the characteristics are reported in the last column in Table 2-1. If high *ECIN* tends to pick up small firms and low *ECIN* tends to pick up large firms, the zero-investment portfolios defined in this way will capture the hedge return related to *ECIN* while eliminating the effect of firm size. Panel C reports the time-series mean of the value-weighted average returns for each portfolio. The average return difference of between the high and low *ECIN* portfolios across all size deciles is 0.433% (*t*-statistics=3.45).

Table 2-2 reports the correlations in returns across the ten *ECIN* portfolios. There are significant positive correlations between each pair of *ECIN* portfolios, except for some correlations involving the largest size decile. The positive correlations between these portfolios suggest the common variation of returns within high *ECIN* stocks and within low *ECIN* stocks, indicating that *ECIN* might capture some underlying systematic risk. To test this effect, we construct the *ECIN* factor, which is the average return of the ten *ECIN*-based zero-investment portfolios in Table 2-1.

⁸ The results are robust if we construct the *ECIN* portfolios and factor as (High *ECIN*- Low *ECIN*) or if we use equal-weighted portfolio returns.

Table 2-3 Panel A gives the summary statistics of the *ECINF* factor (*ECINF*) and the traditional factors, namely, market, size and book-to-market factors of Fama and French (1993) and the momentum factor of Carhart (1997). *ECINF* has a mean of 0.433% and a median of 0.473%. Panel B gives the correlation between factors. *ECINF* is negatively correlated with market and size factors and positively correlated with book-to-market and momentum factors.⁹ The strong correlations between *ECINF* and other factors raise the question whether the returns of *ECINF* can be captured by other risk factors. We run time-series regressions with *ECINF* as the dependent variable and the other factors as explanatory variables. The coefficients are reported in Panel C. The risk adjusted returns of *ECINF* are 0.595% (t -statistics=5.92) in the three-factor regression and 0.407% (t -statistics=4.52) in the four-factor regression. So we conclude that the market, size, book-to-market and momentum factors cannot explain the return variation captured by *ECINF*.

To further test whether *ECINF* is a priced risk factor, we conduct the two-stage cross-sectional regression (2SCSR) at the portfolio level, with controlling for three Fama and French (1993) factors (the market risk premium (MKT), size (SMB), and book-to-market (HML)) and Carhart (1997) momentum factor (UMD).

The first stage of the test is the contemporaneous time-series regression of a single portfolio return on the risk factors, where we get the estimated multivariate loadings (β 's) on the risk factors for each portfolio. For example, when we add

⁹ The correlations between *ECINF* and other factors are -0.622, -0.382, 0.372, and 0.393 respectively for market, size, book-to-market, and momentum factors, if we use the hedge return of (High *ECIN*-Low *ECIN*).

ECINF to the Fama-French three-factor (MKT, SMB and HML) model, the factor loadings (β 's) for portfolio p are estimated from the following time-series regression:

$$R_{p,t} - R_{F,t} = \beta_0 + \beta_{p,MKT}MKT_t + \beta_{p,SMB}SMB_t + \beta_{p,HML}HML_t + \beta_{p,ECINF}ECINF_t + \varepsilon_{p,t} \quad (2.1)$$

The second stage of the test is to run the cross-sectional regression of portfolios' excess returns on the factor betas estimated from equation (2.1).

$$E[R_{p,t} - R_{F,t}] = \gamma_{0,t} + \gamma_{1,t}\beta_{p,MKT} + \gamma_{2,t}\beta_{p,SMB} + \gamma_{3,t}\beta_{p,HML} + \gamma_{4,t}\beta_{p,ECINF} \quad (2.2)$$

If *ECINF* is a priced factor, the risk premium γ_4 should be significantly positive. The cross-sectional regression of equation (2.2) adopts the method same as Black, Jensen and Scholes (1972) and Fama and MacBeth (1973) to mitigate the problem of cross-sectional dependence in data. For each month, we run the cross-sectional regression of portfolios' returns on the estimated factor loadings to get a time-series estimation of γ 's, of which the time-series mean and t -statistic are calculated. Since the factor betas in the second-stage regression are estimates rather than true values, the regression (2.2) is subject to the "errors-in-variable" problem. We use the weighted least-squares procedure outlined by Shanken (1992) to correct standard errors. The adjusted R^2 is the average across all cross-sectional regressions. In addition, we conduct the Gibbons, Ross and Shanken (1989) test in the first stage regressions to investigate whether the estimated intercepts for all portfolios are jointly zero.

We perform the 2SCSR method on three sets of portfolios. The first set consists of 25 portfolios sorted independently by firm size and *ECIN*, including all NYSE/Amex common stocks with valid market equity and *ECIN* data. The portfolios are formed at the end of each December. The size breakpoint is the NYSE market equity quintile at the end of the year. And stocks are independently sorted into quintiles by *ECIN* estimated through the year. The 25 portfolios are the intersection of the five size portfolios and five *ECIN* portfolios. The value-weighted monthly returns during the next 12 months are linked across years to form a single return series for each portfolio.

The second set consists of 27 portfolios sorted independently by firm size, book-to-market (BM) ratio, and *ECIN*, including all NYSE/Amex common stocks with valid market equity, book equity and *ECIN* data. The portfolios are formed at the end of each June. The size breakpoints are the NYSE market equity 0.7 and 0.3 fractiles at the end of each June. BM for each June is the book equity at the end of the last fiscal year divided by the market equity at the end of the last calendar year. The BM breakpoints are the 0.7 and 0.3 fractiles of NYSE book-to-market ratio. The *ECIN* breakpoints for each June are the *ECIN* terciles estimated through last calendar year. As the *ECIN* data is available from the year 1983, the portfolio is formed from July 1984. The value-weighted monthly returns from July of each year through June of next year are linked across years to form a single return series for each portfolio.

The third set consists of the 25 Fama-French portfolios formed by firm size and book-to-market (BM) quintiles. The portfolios, including all NYSE/Amex/Nasdaq stocks, are formed at the end of each June. The size breakpoints are the NYSE market equity quintiles at the end of each June. BM is the book equity at the end of the last

fiscal year divided by the market equity at the end of the last calendar year. The BM breakpoints are NYSE quintiles. The value-weighted monthly returns from July of each year through June of next year are linked across years to form a single return series for each portfolio. As *ECINF* is constructed on NYSE/Amex stocks only, the third set of portfolios gives us the out-of-sample test to illustrate that *ECINF* is a common risk factor across stock markets.

We conduct the two-pass tests on the three-factor (MKT, SMB, and HML) and the four-factor model (MKT, SMB, HML, and UMD), both of which are augmented with *ECINF*. Table 2-4 presents summary results of the first-stage time-series estimation of regression (2.1). We report the *F*-value and *p*-value of Gibbons, Ross and Shanken (1989) test (GRS) and the cross product of regression intercepts (pricing error of portfolios). Adding *ECINF* reduce the *F*-values of GRS test and reduce the cross product of portfolio intercepts. For both the size/*ECIN* portfolios and the size/BM/*ECIN* portfolios, GRS tests reject the null hypothesis of zero intercepts at the 5% significance level for the Fama-French three-factor models, while the tests cannot reject the null hypothesis when *ECINF* is added. Tests on the Carhart four-factor model also support the reduced significance in rejecting the zero intercepts hypothesis when *ECINF* is added. For the third set of portfolios in Panel C, although adding *ECINF* into the factor models also reduce *F*-values, the GRS test still significantly rejects the null hypotheses of zero intercepts. A closer investigation of this set of portfolios reveals that the abnormal returns mostly come from small or low book-to-market portfolios and Nasdaq stocks. The cross products of pricing errors are economically small.

Table 2-5 reports the coefficients of the second stage cross-sectional regressions. Panel A represents the results of the Fama-French three-factor model and Panel B represents the results of the Fama-French three-factor model augmented by *ECINF*. Comparing Panel A and Panel B, we can see that adding *ECINF* into the 2SCSR increases the adjusted R^2 for all three sets of portfolios.¹⁰ In addition, Panel B shows that the loadings on *ECINF* have significant explanation power for the cross-sectional variation of the three sets of portfolio returns. The coefficients on the loading of *ECINF* for the three sets of portfolios are significant and positive at 0.750 (t -statistics =2.73), 0.676 (t -statistics=3.19), and 1.058 (t -statistics=4.96) respectively. This empirical evidence supports the hypothesis that *ECINF* is a priced risk factor. It is consistent with Easley and O'Hara (2004) that information risk is a systematic risk and the investors require positive premium for holding this risk. To further test whether the significance of *ECINF* loading is due to the omission of momentum factor, we add the momentum factor *UMD* into the test and report the results in Panel C and Panel D. *ECINF* loadings are significant and positive in the five-factor model (Panel D) at 0.507, 0.608, and 1.071 (t -statistics are 4.12, 2.75, and 5.10 respectively). The adjusted R^2 is higher in Panel D than in Panel C. Overall, the two-stage cross-sectional regression tests provide evidence that the *ECIN* is a priced risk factor.

¹⁰ The insignificant (or negative) coefficients on the market and size factors loadings may be due to the fact that beta and size effects are disappearing in the asset pricing around the recent sample period. Mohanram and Rajgopal (2008) and Core, Guay and Verdi (2008) find similar negative or insignificant coefficients for the market and size factors loadings in the comparable sample periods.

2.4 Asset Pricing Anomalies

In this section, we exploit *ECINF* in order to explain the financial anomalies, namely price momentum, post-earnings announcement drift, underperformance in long-run stock returns following SEO, and the relations between equity issuance and the long-run stock returns. Financial anomalies violate the informational efficient market assumption as stock prices do not adjust immediately to reflect all information. Since financial anomalies are mostly following public news events, long-term drift is puzzling under the assumptions of rational investors and informational efficient market. One possible explanation for price momentum and earnings momentum could be related to variations in interpreting public information. Though the performances of the past returns or news of quarterly earnings are publicly observable, investors may interpret and react to such information differently. As a result, public information may have private components. Price momentum and earnings momentum could reflect the slow diffusion of such private information. On the other hand, SEO underperformance as well as issuance puzzles may reveal reduced information risk of issue firms. For example, firms may time SEO when their information risk is suddenly reduced. During the SEO process, firm disclosures more information to market and the outsiders has more incentive to explore news about the SEO firm. As we have shown in the previous section, information risk is a priced fundamental risk. If the underperformance of SEO or the equity issuance puzzle is related to the reduced information risk, we expect that the information risk factor can fully capture any abnormal returns that cannot be explained by conventional risk factors.

Fama (1998) noted that anomalies in long-term post-event returns can be over-emphasized when event firms are equal-weighted because of the bad-model problem in small size firms. To avoid this problem, we focus our analysis on value-weighted portfolio returns, which would also give a more accurate assessment on the total wealth effects of investors. Furthermore, we choose the average monthly abnormal return rather than the buy-and-hold abnormal return, as this method gives clear assessment on the contribution of *ECINF* to a traditional risk factors model. In addition, the average monthly abnormal return poses fewer statistical problems (non-normal and skewness), corrects the cross-correlation problem in long-run returns, and has the same return period as assumed by most asset pricing models (Fama (1998).)

2.4.1 Price Momentum

We define the momentum strategies following those proposed by Jegadeesh and Titman (1993). At the beginning of each month stocks are sorted into quintiles based on the past J months' return skipping the last one month. Firms in the top (bottom) quintile are called "winners" ("losers"). The portfolio return is the value-weighted average of component stocks' returns. Each portfolio is held for K months and liquidated after that. Therefore, in each month we have K portfolios in each momentum quintile that are formed in the current month as well as in the previous $K-1$ months. Each portfolio return is weighted by $1/K$. The zero-investment portfolio has a long (short) position on the winners (losers). The average monthly abnormal return is the intercept, α_p , of the time-series regression

$$R_{pt} - R_{ft} = \alpha_p + \beta_p [R_{Mt} - R_{ft}] + s_p SMB_t + h_p HML_t + \varepsilon_{pt} \quad (2.3)$$

or

$$R_{pt} - R_{ft} = \alpha_p + \beta_p [R_{Mt} - R_{ft}] + s_p SMB_t + h_p HML_t + e_p ECINF_t + \varepsilon_{pt} \quad (2.4)$$

where R_{pt} is the monthly average return of momentum portfolios; R_{ft} is the one-month Treasury bill rate; R_{Mt} is the monthly return on the value-weighted market portfolio of NYSE, Amex, and Nasdaq stocks; SMB_t and HML_t are size and book-to-market factors devised by Fama and French (1993) respectively; $ECINF_t$ is information risk factor constructed in Section 3.

Table 2-6 reports the intercepts of the three-factor regression (regression (2.3),) and the intercepts and the $ECINF$ loadings of the four-factor regression (regression (2.4).) Momentum portfolios are formed on NYSE/Amex firms with different trading strategies ($J=6, 12$ and $K=6, 12$). The last row reports the regression results for the zero-investment portfolios. We observe increasing $ECINF$ loadings from losers to winners. The positive factor loadings of the winners reflect positive shocks to the market-wide information asymmetry environment; the negative factor loadings of the losers suggest that losers can be a hedge against information risk. Different information risk exposures may be related to the intensive informed trading in winners and the dominating noise trading in losers during the portfolio formation period. Because of short-sale costs and constraints, informed speculation in losers is limited and sometimes impossible. In addition, we observe that the abnormal returns of momentum strategies are largely diminished when we control for $ECINF$. For

example, the zero-investment portfolio of (6, 6) strategy has a positive Fama-French risk-adjusted abnormal return 0.669 (t -statistics=3.03). The abnormal return is reduced to an insignificant 0.133 (t -statistics=0.64) in the four-factor model. The loading on *ECINF* is significantly positive at 0.956(t -statistics=8.17). Table 2-7 reports the results of the extended tests on the NYSE/Amex/Nasdaq firms. The regression results are similar to those of the NYSE/Amex firms, if somewhat weaker for the winners, perhaps due to the noise in explaining Nasdaq firms with *ECINF* constructed from NYSE/Amex firms. Therefore, the abnormal returns in the momentum strategies can be an information risk premium since winners are exposed to more information risk than are losers.

2.4.2 Post-Earnings-Announcement Drift

We calculate the quarterly standardized earnings surprises (SUE) with three different specifications. The first specification is the traditional method based on the seasonal random walk (SRW) model following Chan, Jegadeesh and Lakonishok (1996) and Chordia and Shivakumar (2006). It is estimated as

$$SUE_{i,q} = \frac{(E_{i,q} - E_{i,q-4})}{\sigma_{i,q}} \quad (2.5)$$

where $E_{i,q}$ is primary Earnings Per Share (EPS) before extraordinary items in quarter q for stock i , and $\sigma_{i,q}$ is the standard deviation of earnings changes in the prior eight quarters. $E_{i,q}$ is unadjusted for stock splits, but $E_{i,q-4}$ is adjusted for any stock splits and stock dividends during the period $(t-4, t)$. Following Chordia and Shivakumar (2006) and Sadka (2006), the earnings surprise is scaled by the standard deviation rather than

by the stock price, total asset, or sales since these variables may proxy for size or expected returns, therefore preventing the introduction of bias into the tests. Livnat and Mendenhall (2006) suggest that omitting special items from *EPS* may change the profitability of traditional SUE analysis and this method is closer to the analysts' or the street's measure. Accordingly, the second measure of SUE is calculated after such exclusion. In particular, 65% times the amount of special items divided by the number of shares, the result of which is subtracted from the *EPS* in equation (2.5). The third measure of SUE is the difference between actual earnings and analysts' forecasts; that is, we replace the SRW forecast ($E_{i,q-4}$) in equation (2.5) with a measure of analysts' expectations, which is the median of analysts' forecasts reported to I/B/E/S in the 90 days prior to the earnings announcement.

At the beginning of each month t , the SUE portfolios are formed based on the most recent earnings announcement which is not in the current month t and also no earlier than four months before the formation date. We sorted firms into quintiles based on the SUE in each month. Stock returns are value-weighted in each portfolio. The positions are held for three months starting from the formation month¹¹. Therefore, in each month, we have three portfolios in each SUE quintile that are formed in the current month as well as in the previous two months, and each portfolio return is weighted by 1/3. The zero-investment portfolio takes a long (short) position on the highest (lowest) SUE quintile. The portfolios' returns are run on the following

¹¹ Results are robust if we use six months holding period.

risk-factor models and the average monthly abnormal return is the intercept, α_p , of the time-series regressions

$$R_{pt} - R_{ft} = \alpha_p + \beta_p [R_{Mt} - R_{ft}] + s_p SMB_t + h_p HML_t + \varepsilon_{pt} \quad (2.6)$$

or

$$R_{pt} - R_{ft} = \alpha_p + \beta_p [R_{Mt} - R_{ft}] + s_p SMB_t + h_p HML_t + e_p ECINF_t + \varepsilon_{pt} \quad (2.7)$$

where R_{pt} is the monthly average return of the SUE portfolios; R_{ft} is the one-month Treasury bill rate; R_{Mt} is the monthly return on the value-weighted market portfolio of NYSE, Amex, and Nasdaq stocks; SMB_t and HML_t are size and book-to-market factors devised by Fama and French (1993) respectively; $ECINF_t$ is the information risk factor constructed in Section 3.

Table 2-8 reports the intercepts of the three-factor regression (regression (2.6),) and the intercepts and the $ECINF$ loadings of the four-factor regression (regression (2.7).) SUE portfolios are formed on NYSE/Amex firms. As observed for price momentum, we see the increasing loading on $ECINF$ from negative to positive as moving up SUE quintile. Similar to the losers in the price momentums, the bad news firms can be a hedge to information risk because informed trading is less likely in these firms compared to in good news firms due to short-sale costs and constraints. The loadings of zero-investment portfolios on $ECINF$ are significantly positive at 0.582, 0.716 and 0.306 for the three SUE measures respectively. The monthly abnormal returns of the zero-investment portfolios in the three-factor models are positive and significant at 0.37% to 0.43% for the three SUE measures. However, the risk-adjusted return is not significant when $ECINF$ is added. Livnat and Mendenhall

(2006) argue that SUE formed on the analysts' forecasts predict higher PEAD, which may capture different kinds of mispricing than the SRW model. Our results show that the PEAD based on analysts' forecasts and the SRW model can both be explained by the information risk factor, while the explanation power of *ECINF* is somewhat weaker in the high SUE quintiles formed on analyst's forecast. We then include Nasdaq firms into the sample for extended tests. The results are listed in Table 2-9, and are similar to those using NYSE/Amex sample. As we argued earlier, the empirical results suggest that the PEAD reflects different exposures to information risk caused by asymmetric informed trading between good news and bad news firms. Therefore the lower returns in the low SUE stocks can be explained by lower loadings on information risk and accordingly lower information risk premium.

2.4.3 Long-Run Underperformance after SEO and Stock Issuance

Following Loughran and Ritter (1995), Brav and Gompers (1997), Brav, Geczy and Gompers (2000) and Shivakumar (2000), our inference on SEO underperformance is based on monthly average returns. Fama (1998) suggests that this calendar-time approach corrects for the correlation of returns across events and is not subject to the problem of skewness in long-horizon buy-and-hold returns. Figure 2-1 plots the distribution of SEOs in each year from 1979 to 2007 for exchanges NYSE/Amex/Nasdaq. Following Loughran and Ritter (1995) and Spiess and Affleck-Graves (1995), we document the abnormal return in both three and five years post-SEO periods. Every month from January 1984 to December 2007, we assign firms

that have finished SEO in the previous 36 months (60 months) into the three-year (five-year) holding portfolio. The value-weighted returns¹² are run on the contemporaneous Fama-French and *ECIN* risk factors to assess the portfolio's abnormal returns. The average monthly abnormal return during the three-year (five-year) post-event period is the intercept, α_p , of the time-series regressions

$$R_{pt} - R_{ft} = \alpha_p + \beta_p [R_{Mt} - R_{ft}] + s_p SMB_t + h_p HML_t + \varepsilon_{pt} \quad (2.8)$$

or

$$R_{pt} - R_{ft} = \alpha_p + \beta_p [R_{Mt} - R_{ft}] + s_p SMB_t + h_p HML_t + e_p ECINF_t + \varepsilon_{pt} \quad (2.9)$$

where R_{pt} is the monthly return on the SEO portfolios; R_{ft} is the one-month Treasury bill rate; R_{Mt} is the monthly return on the value-weighted market portfolio of NYSE, Amex, and Nasdaq stocks; SMB_t and HML_t are size and book-to-market factors devised by Fama and French (1993) respectively; $ECINF_t$ is information risk factor constructed in Section 3.

We report the changes in *ECIN* around the SEO years and the time series regressions of the SEO portfolio in Table 2-10. Panel A reports the changes of *ECIN* around the SEO year. We calculate the yearly average of *ECIN* in the pre-SEO year

¹² For stocks making SEO in year t , the post-SEO monthly returns are all weighted by market capitalization at the end of year $t-1$, in order to avoid the effects of the equity issues on the weights. We also do robustness tests by using the market capitalization at the end of each year as the weight for the following year's monthly return. The results are similar to those reported in this paper.

and in three or five post-SEO years respectively. The SEO year is not included so that any informed speculating around SEO will not affect the results. The difference between the *ECIN* in the three (five) post-SEO years and the *ECIN* in the pre-SEO year is significantly negative at -0.055 (-0.072). We further run the time-series regressions (regression (2.8) and regression (2.9)) of the pre-SEO and post-SEO portfolios. In Panel B of Table 2-10 we observe that the pre-SEO portfolio does not have significant loading on *ECINF*. However, the post-SEO portfolios have significantly negative loadings on *ECINF* and the negative loadings on *ECINF* can explain the post-SEO abnormal return. The three-year post-SEO portfolio generates a significantly abnormal return of -0.330% (t -statistics=-2.59) adjusted to the Fama-French risk factors. However, when we include *ECINF* into the regression, the risk-adjusted return becomes insignificantly different from zero. In addition, the significantly negative loading (-0.221 with t -statistics=-2.96) on *ECINF* suggests a negative information risk exposure of post-SEO firms. Similar results can be observed for the five-year post-SEO portfolio. The negative loadings on the information risk factor and the reduced *ECIN* characteristic in the post-SEO years suggest that the long-run underperformance of SEO firms can be a manifestation of reduced information risk. This is consistent with the argument that firms reveal information to the public during SEO and therefore reduce information asymmetry risk afterwards. In addition, the fact that information risk can explain the puzzle of SEO may also suggest that (at least some) firms time SEO when their information risk is suddenly reduced. We verify this argument by directly running VECM for SEO firms in the year (12 months) immediately before SEO. We find comparing to the first 9 months, *ECIN* significantly drops by 0.023 (t -statistics= -5.55) in the last three month before SEO.

Recently, Pontiff and Woodgate (2008) and Fama and French (2008) document the cross-sectional correlation between share issuance and stock returns. That is, the stocks with increases in the adjusted shares outstanding (positive issuance) underperform the stocks with decreases in the adjusted shares outstanding (negative issuance). We would investigate whether this issuance effect is related to information risk and whether the difference in the information risk can capture the variation in stock returns. We follow the same procedures as do Pontiff and Woodgate (2008) in calculating the share issue and form the portfolios accordingly. In the time-series regression, we investigate whether the information risk factor *ECINF* helps to explain the variation in returns of share issue portfolios. Specifically, we first obtain the number of shares outstanding and the factor to adjust shares outstanding from CRSP. We then calculate the cumulative product of the CRSP adjustment factor f as the total factor at time t :

$$Total\ Factor_t = \prod_{i=1}^t (1 + f_i) \quad (2.10)$$

The number of shares outstanding adjusted for splits and other events is

$$AdjustedShares_t = SharesOutstanding_t / TotalFactor_t \quad (2.11)$$

Therefore, the annual share issue at time t is

$$ISSUE_{t,t-11} = Ln(AdjustedShares_t) - Ln(AdjustedShares_{t-11}) \quad (2.12)$$

Following Pontiff and Woodgate (2008), we require a gap of six months between the portfolio formation month and the beginning month of the portfolio holding period to account for the possible timing mismatch between shares

outstanding and CRSP-reported shares outstanding. That is, at the beginning of each month, we form the portfolio according to the sign of $ISSUE_{17,-6}$ and hold the portfolio for three years to observe the long-run return. The portfolio return is the value-weighted average of the component stock returns. In each month we have the zero-investment portfolio that takes long (short) positions on the positive (negative) $ISSUE_{17,-6}$ portfolios formed in the current and past 35 months, and each portfolio is weighted by $1/36$.¹³ Finance literature¹⁴ documents that firms announce stock repurchase to signal good information. This signaling effect can reduce information asymmetry and therefore decrease information risk. To eliminate such signaling effect associated with stock repurchase, we deleted firms with repurchase announcements in the SDC dataset. The portfolio is rebalanced every month and the return is regressed on the contemporaneous Fama-French and $ECIN$ risk factors in order to assess the portfolio's abnormal returns. The average monthly abnormal return during the three-year post-event period is the intercept, α_p , of the time-series regressions

$$R_{pt} - R_{ft} = \alpha_p + \beta_p [R_{Mt} - R_{ft}] + s_p SMB_t + h_p HML_t + \varepsilon_{pt} \quad (2.13)$$

or

¹³ The untabulated results for the zero-investment portfolios with a long (short) position on the positive (zero and negative) $ISSUE_{17,-6}$ portfolios are similar to the results presented in this paper.

¹⁴For example, Dann (1981), Masulis (1980), Comment and Jarrell (1991), Ofer and Thakor (1987), Hausch and Seward (1993), and Vermaelen (1981).

$$R_{pt} - R_{ft} = \alpha_p + \beta_p [R_{Mt} - R_{ft}] + s_p SMB_t + h_p HML_t + e_p ECINF_t + \varepsilon_{pt} \quad (2.14)$$

where R_{pt} is the monthly return on the *ISSUE* portfolios; R_{ft} is the one-month Treasury bill rate; R_{Mt} is the monthly return on the value-weighted market portfolio of NYSE, Amex, and Nasdaq stocks; SMB_t and HML_t are size and book-to-market factors devised by Fama and French (1993) respectively; $ECINF_t$ is information risk factor constructed in Section 3.

Table 2-11 reports the coefficients of the time-series regressions of the zero-investment portfolios. The abnormal return is significantly negative in the three-factor model at -0.205 (t -statistic=-2.33). Adding *ECINF* into the regression reduce the abnormal return to an insignificant level. The negative loading on *ECINF* (-0.163, t -statistic=-3.47) suggests lower information risk exposure of the positive issue firms than of the negative issue firms. As the long-run abnormal return and the negative information risk loading of the positive issue firms is strongly related to the performance of the SEO firms demonstrated previously, we explore whether removing the data associated with SEO affects our results. Firms with SEO announcement in the SDC database in the last three years are deleted from the sample. The zero-investment portfolio has a significant negative return at -0.183(t -statistic=-2.10) in the Fama-French three-factor regression. The regression on the four-factor model yields an insignificant intercept and a negative loading on *ECINF* at -0.153 (t -statistic=-3.28). This significant negative loading on *ECINF* reveals that the issuance effect is related to information risk. The underperformance of positive issue firms relative to the performance of negative issue firms reflects the lower information risk premium.

2.5 Conclusion

In this paper, we create the information risk factor *ECINF* based on the zero-investment portfolio, which is size neutral but takes a long position on high *ECIN* stocks and a short position on low *ECIN* stocks. The average return of *ECINF* is significant and positive, and cannot be explained by Fama and French (1993) risk factors and Carhart (1997) momentum factor. The two-stage cross-sectional asset pricing tests on three different sets of portfolios reveal significantly positive risk premiums on *ECINF*, which can be viewed as explicit evidence supporting Easley and O'Hara (2004)'s argument that information risk is a priced systematic risk. The significant risk premiums on *ECINF* is a sharp contrast to the findings in Mohanram and Rajgopal (2008) that the loadings of *PIN* factor are insignificantly related to the cross-sectional returns, suggesting that *ECINF* captures the underlying information risk more efficiently than the *PIN* factor.

In addition, *ECINF* helps explain the financial anomalies. Because most of the anomalies are news-related, we expect their associated returns to be sensitive to the shocks of the market-wide information risk factor. The zero-investment portfolios based on price and earnings momentum have significant positive loadings on *ECINF*. Also, the abnormal returns are not significant when *ECINF* is included into the Fama-French three-factor regressions. This supports the informational risk explanation for the financial anomalies that the underreaction to past returns or to earnings surprises can be explained as the difference in information risk between winners (high SUE) and losers (low SUE) firms. The negative loadings on the bad news portfolios reflect that these portfolios are a hedge against fluctuation in market-wide information risk. It

also indicates that there are fewer information-based trades in bad news firms than the average firms possibly due to short-sale constraints.

In addition, we employ *ECINF* to explain the long-run underperformance of the SEO stocks. The post-SEO firms have significantly negative loadings on information risk, and their negative abnormal returns can be fully captured by *ECINF*. This suggests that the post-SEO underperformance can be a manifestation of the reduced information risk for the issuing firms, because SEO provides extra public information and increases firms' information dissemination. Similarly, we conduct these tests on positive and negative shares issuance firms. The zero-investment portfolios with a long (short) position on the positive (negative) shares issuance firms generate a significant Fama-French risk adjusted return. But the abnormal return is reduced to an insignificant level when we include *ECINF* into the regression. The portfolio's negative loading on *ECINF* reveals that the underperformance of positive issuance firms reflects the lower information risk of these firms compared to those of the negative issuance firms. Furthermore, we illustrate that the share issuance effect is not driven by the SEO firms.

This paper not only provides evidence that information asymmetry risk is a priced risk factor, but also shows that the most consistent asset pricing anomalies, namely price momentum, PEAD and the share issuance effect, can be explained by information risk. As financial anomalies are news-related and investors are different in both opinion and ability to process material public information, ignoring information risk or returns' sensitivity to the shocks of the market-wide information asymmetry raises the unexplainable abnormal returns. The traditional asset pricing model cannot explain the financial anomalies since they largely assume symmetric

information and completely ignore information risk. However, the empirical results in this paper illustrate the need for asset-pricing models that incorporate information risk. This essay, together with the first essay, suggests that information risk should be considered in asset pricing tests. Failure to do so can produce misleading anomalies that are then interpreted as evidence against market efficiency.

Table 2-1 Characteristics and Returns of *ECIN* Portfolios

This table reports the characteristics and returns of the *ECIN* portfolios. At the end of each year, we sort all NYSE/Amex stocks with non-missing year-end market capitalization and yearly *ECIN* data into size deciles. Within each size decile, three portfolios are sorted by *ECIN*s and three portfolios are sorted by firm size, where the two sorts are independent. The hedge return of the *ECIN* portfolio for each size decile is the value-weighted monthly return of (High *ECIN*- Small Size)-(Low *ECIN*- Large Size) in the following year. Panel A reports the average *ECIN*. Panel B reports the average market capitalization (in millions). Panel C reports the percentage value-weighted monthly return in percentage. Rows represent the size deciles in the first sort. Columns represent the second *ECIN* or size sort. *LEGIN* denotes the lowest *ECIN* tercile, *MEGIN* denotes the medium *ECIN* tercile, and *HEGIN* denotes the highest *ECIN* tercile. *SSIZE* denotes the smallest size tercile, *MSIZE* denotes the medium size tercile, and *LSIZE* denotes the largest size tercile. The last column, *ECIN* portfolio reports the characteristics of portfolio (High *ECIN*- Small Size)-(Low *ECIN*- Large Size). We take the cross-sectional mean of the characteristics for each portfolio and report the time-series average. The sample period is from January 1983 through December 2006 for *ECIN*s and sizes, and from January 1984 through December 2007 for stock returns.

Panel A: *ECIN*

Size Deciles	LEGIN	MEGIN	HEGIN	SSIZE	MSIZE	LSIZE	ECIN Portfolio
Small	-0.133	-0.076	-0.040	-0.075	-0.083	-0.089	0.078
2	-0.196	-0.106	-0.061	-0.105	-0.120	-0.136	0.103
3	-0.246	-0.148	-0.085	-0.145	-0.160	-0.173	0.133
4	-0.287	-0.182	-0.110	-0.187	-0.190	-0.201	0.163
5	-0.322	-0.207	-0.121	-0.206	-0.218	-0.225	0.183
6	-0.367	-0.247	-0.152	-0.253	-0.250	-0.261	0.207
7	-0.409	-0.281	-0.171	-0.271	-0.294	-0.296	0.212
8	-0.448	-0.321	-0.198	-0.312	-0.324	-0.331	0.232
9	-0.491	-0.353	-0.219	-0.342	-0.353	-0.366	0.248
Large	-0.550	-0.383	-0.224	-0.383	-0.386	-0.392	0.317
Average	-0.345	-0.230	-0.138	-0.228	-0.238	-0.247	0.188

Panel B: Market Capitalization

Size Deciles	LEGIN	MEGIN	HEGIN	SSIZE	MSIZE	LSIZE	ECIN Portfolio
Small	44	42	40	19	35	54	31
2	115	108	104	76	103	133	46
3	223	213	206	169	207	251	66
4	358	358	349	296	347	404	98
5	564	553	544	467	541	629	143
6	863	872	854	728	842	981	244
7	1,419	1,369	1,364	1,144	1,344	1,590	392
8	2,432	2,384	2,354	1,891	2,271	2,822	853
9	5,145	5,076	4,944	3,610	4,658	6,199	2,389
Large	55,479	53,395	56,346	8,655	14,447	78,024	70,236
Average	6,664	6,437	6,711	1,705	2,480	9,109	7,450

Panel C: Stock Returns (%)

Size Deciles	LEGIN	MEGIN	HEGIN	SSIZE	MSIZE	LSIZE	ECIN Portfolio	t-value
Small	-0.070	0.730	1.096	0.369	0.494	0.670	1.467	4.34
2	0.503	0.863	1.021	0.670	0.757	0.888	0.737	2.79
3	0.782	1.200	1.097	1.125	1.041	0.949	0.140	0.50
4	0.652	1.190	1.418	1.010	1.078	1.146	0.902	3.96
5	1.005	1.038	1.272	1.153	1.040	1.118	0.232	1.07
6	1.172	1.170	1.320	1.202	1.286	1.173	0.120	0.52
7	1.015	1.231	1.279	1.136	1.209	1.159	0.287	1.56
8	1.223	1.284	1.283	1.248	1.280	1.250	0.062	0.36
9	1.239	1.206	1.265	1.093	1.236	1.320	0.254	1.63
Large	1.051	1.156	1.160	1.107	1.081	1.128	0.131	0.71
Avg	0.857	1.107	1.221	1.011	1.050	1.080	0.433	3.45

Table 2-3 Descriptive Statistics and Time-Series Relations of Risk Factors

This table reports the descriptive statistics and time-series relations of the three Fama and French (1993) factors, Carhart (1997) momentum factor and the *ECINF* factor. *MKT*, the excess return on the market, is the value-weighted return on all NYSE/Amex/Nasdaq stocks minus the one-month Treasury bill rate. *SMB* (Small minus Big) is the return to size factor-mimicking portfolio. *HML* (High minus Low) is the return to book-to-market factor-mimicking portfolio. *UMD* is the return to momentum factor-mimicking portfolio. *ECINF* is the return to information risk factor-mimicking portfolio. Panel A reports the descriptive statistics of the factors. Panel B reports the correlations. Panel C report the time series regression of *ECINF* on the three/four other risk factors. The sample period is from January 1984 to December 2007.

Panel A: Descriptive Statistics

	N	Mean	STD	MIN	Median	MAX	<i>t-stat</i>
MKT	288	0.636	4.283	-23.130	1.045	12.430	2.52
SMB	288	0.004	3.299	-16.790	-0.190	21.960	0.02
HML	288	0.361	3.097	-12.400	0.315	13.850	1.98
UMD	288	0.822	4.262	-25.060	0.955	18.390	3.27
ECINF	288	0.433	2.130	-13.123	0.473	6.734	3.45

Panel B: Correlation Matrix of the Factors

	MKT	SMB	HML	UMD
SMB	0.200			
HML	-0.488	-0.414		
UMD	-0.099	0.100	-0.063	
ECINF	-0.512	-0.475	0.332	0.379

Panel C: Time-Series Regressions

Regression	Intercept	MKT	SMB	HML	UMD	Adj. R^2
(1)	0.595 (5.92)	-0.229 (-8.81)	-0.264 (-8.15)	-0.043 (-1.11)	0.006 (5.92)	0.403
(2)	0.407 (4.52)	-0.197 (-8.56)	-0.286 (-10.08)	-0.014 (-0.42)	0.192 (9.46)	0.545

**Table 2-4 Portfolio Time-Series Regressions of Contemporaneous Excess Returns on
Factor Returns**

This table reports the average loadings and average adjusted R^2 of portfolio excess returns on the contemporaneous factor returns in the time-series regressions. The first set of 25 portfolios is formed on firm size and *ECIN* quintiles. The second set of 27 portfolios is formed on firm size, book-to-market (BM) ratio, and *ECIN* terciles. The third set of portfolios is the Fama-French 25 portfolios formed on size and book-to-market quintiles. *MKT*, the excess return on the market portfolio, is the value-weighted return on all NYSE/Amex/Nasdaq stocks minus the one-month Treasury bill rate. *SMB* (Small minus Big) is the return to size factor-mimicking portfolio. *HML* (High minus Low) is the return to book-to-market factor-mimicking portfolio. *UMD* is the return to momentum factor-mimicking portfolio. *ECINF* is the return to information risk factor-mimicking portfolio. GRS F -value is the F -statistic of Gibbons, Ross and Shanken (1989), $F = \left(\frac{T}{N}\right) \left(\frac{T-N-K}{T-K-1}\right) \left(\frac{\hat{\alpha}' \hat{\Sigma}^{-1} \hat{\alpha}}{1 + \hat{\mu}' \hat{\Omega}^{-1} \hat{\mu}}\right)$, where T is the number of observations, K is the number of factors, N is the number of portfolios, $\hat{\alpha}$ is the $N \times 1$ vector of estimated intercepts, $\hat{\Sigma}$ is the unbiased estimate of the residual covariance matrix, $\hat{\mu}$ is the $K \times 1$ vector of the factor portfolios' sample means, and $\hat{\Omega}$ is the unbiased estimate of the factor portfolios' covariance matrix. This statistic has an F -distribution with N and $(T-N-K)$ degrees of freedom. If $\hat{\alpha}_i = 0 \forall i$, then the GRS F -statistic equals zero; the larger the α 's are in absolute value the greater the GRS F -statistic. We also report the pricing error as $INT = \sqrt{\hat{\alpha}' \hat{\alpha} / N}$. The sample period is from January 1984 to December 2007 for 25 size-*ECIN* portfolios and for 25 size-BM portfolios, and is from July 1984 to December 2007 for 27 size-BM-*ECIN* portfolios.

Panel A: 25 Size and *ECIN* ($T=288$)

MKT	SMB	HML	UMD	ECINF	Adj. R^2	INT (%)	GRS <i>F</i> -value	GRS <i>p</i> -value
1.094	0.394	0.434			0.711	0.544	2.185	0.001
1.033	0.316	0.424		-0.274	0.725	0.314	1.244	0.201
1.067	0.406	0.415	-0.163		0.726	0.384	1.673	0.026
1.05	0.378	0.412	-0.145	-0.093	0.735	0.245	1.194	0.245

Panel B: 27 Size, BM, and *ECIN* ($T=282$)

MKT	SMB	HML	UMD	ECINF	Adj. R^2	INT (%)	GRS <i>F</i> -value	GRS <i>p</i> -value
1.096	0.308	0.537			0.714	0.406	1.927	0.005
1.060	0.267	0.53		-0.155	0.720	0.274	1.294	0.157
1.072	0.324	0.516	-0.139		0.726	0.257	1.402	0.095
1.077	0.331	0.516	-0.144	0.025	0.729	0.199	1.012	0.453

Panel C: 25 Size and BM ($T=288$)

MKT	SMB	HML	UMD	ECINF	Adj. R^2	INT (%)	GRS <i>F</i> -value	GRS <i>p</i> -value
1.078	0.518	0.271			0.835	0.217	2.877	0.000
1.032	0.465	0.263		-0.200	0.839	0.149	2.242	0.001
1.050	0.537	0.247	-0.163		0.850	0.169	2.830	0.000
1.051	0.538	0.247	-0.163	0.005	0.851	0.153	2.351	0.001

Table 2-5 Monthly Cross-Sectional Regressions of Returns on Factor Loadings

This table reports the monthly cross-sectional regression of portfolio excess returns on factor loadings. The first set of 25 portfolios is formed on firm size and *ECIN* quintiles. The second set of 27 portfolios is formed on firm size, book-to-market (BM) ratio, and *ECIN* terciles. The third set of portfolios is the Fama-French 25 portfolios formed on size and book-to-market quintiles. The return data of the Fama-French 25 portfolios is obtained from Kenneth French's website. Value weighted monthly excess returns for these portfolios are regressed against factors loadings of the market (*MKT*), size (*SMB*), book-to-market (*HML*), momentum (*UMD*) and *ECIN* (*ECINF*). The cross-sectional regression is estimated for each month in the sample period over year 1984 to 2007. Then the coefficients of cross-sectional regressions are averaged through time and *t*-statistics (in the parentheses) of the time-series means are calculated accordingly. Panel A reports the regressions on the factor loadings from the Fama-French three-factor model (*MKT*, *SMB*, and *HML*). Panel B reports the regressions on the factor loadings from the Fama-French three-factor model (*MKT*, *SMB*, and *HML*) augmented by *ECIN* factor (*ECINF*). Panel C reports the regressions on the factor loadings from the Fama-French four-factor model (*MKT*, *SMB*, *HML*, and *UMD*). Panel D reports the regression on the factor loadings from the Fama-French four-factor model (*MKT*, *SMB*, *HML*, and *UMD*) augmented by *ECIN* factor (*ECINF*).

Panel A Regressions Using Loadings from Fama-French three-factor Model

Portfolios	LMKT3	LSMB3	LHML3	Adj. R2
25 size and <i>ECIN</i>	-1.002 (-3.15)	-0.669 (-3.72)	1.135 (4.07)	0.267
27 size, BM, and <i>ECIN</i>	-1.204 (-3.15)	-0.373 (-1.90)	0.700 (3.28)	0.289
25 size and BM	-1.024 (-2.40)	-0.215 (-1.52)	0.291 (2.26)	0.432

Panel B Regressions Using Loadings from Fama-French three-factor Model with *ECIN* Added

Portfolios	LMKT4	LSMB4	LHML4	LEGIN4	Adj. R2
25 size and <i>ECIN</i>	0.666 (1.37)	-0.126 (-0.23)	-0.337 (-1.69)	0.750 (2.73)	0.286
27 size, BM, and <i>ECIN</i>	-0.271 (-0.48)	-0.259 (-1.38)	0.592 (2.89)	0.676 (3.19)	0.316
25 size and BM	-0.657 (-1.49)	-0.021 (-0.14)	0.238 (1.76)	1.058 (4.96)	0.484

Panel C Regressions Using Loadings from Fama-French four-factor Model

Portfolios	LMKT4	LSMB4	LHML4	LUMD4	Adj. R2
25 size and <i>ECIN</i>	0.022 (0.05)	-0.195 (-0.94)	0.869 (3.29)	2.340 (3.02)	0.312
27 size, BM, and <i>ECIN</i>	-0.919 (-1.64)	-0.333 (-1.59)	0.658 (2.98)	0.449 (0.67)	0.286
25 size and BM	-0.549 (-1.11)	-0.033 (-0.22)	0.252 (1.92)	2.299 (3.13)	0.450

Panel D Regressions Using Loadings from Fama-French four-factor Model with *ECIN* Added

Portfolios	LMKT5	LSMB5	LHML5	LUMD5	LECIN5	Adj. R2
25 size and <i>ECIN</i>	-0.057 (-0.11)	-0.234 (-1.13)	0.907 (3.2)	2.174 (2.54)	0.507 (4.12)	0.314
27 size, BM, and <i>ECIN</i>	-0.421 (-0.73)	-0.287 (-1.45)	0.589 (2.71)	0.510 (0.74)	0.608 (2.75)	0.321
25 size and BM	-0.632 (-1.35)	-0.015 (-0.1)	0.224 (1.64)	1.325 (1.94)	1.071 (5.10)	0.490

Table 2-6 Momentum Portfolios on NYSE/Amex

This table reports the time-series regression of the price momentum portfolios. The sample includes all common stocks in NYSE/Amex for years 1984-2007 with valid CRSP data. The portfolios are formed based on the previous J months returns (skipping the last month) and held for K months. The dependent variables are the excess returns for momentum quintiles or the zero-investment portfolios returns with long positions on winners firms and short positions on losers firms. The value-weighted portfolios returns are regressed on the contemporaneous three Fama-French risk factors (MKT , SMB , and HML) and the information risk factor $ECINF$. MKT , the excess return on the market, is the value-weighted return on all NYSE/Amex/Nasdaq stocks minus the one-month Treasury bill rate. SMB (Small minus Big) is the return to size factor-mimicking portfolio. HML (High minus Low) is the return to book-to-market factor-mimicking portfolio. $ECINF$ is the return to information risk factor-mimicking portfolio. The table reports the alphas of the Fama-French three-factor model, the alphas and the $ECINF$ factor loadings of the Fama-French three-factor model augmented by $ECINF$. The t -statistics are in parentheses.

	J=6, K=6			J=6, K=12			J=12, K=6			J=12, K=12		
	Alpha FF3	Alpha ECINF	ECINF	Alpha FF3	Alpha ECINF	ECINF	Alpha FF3	Alpha ECINF	ECINF	Alpha FF3	Alpha ECINF	ECINF
Loser	-0.463 (-2.98)	-0.062 (-0.43)	-0.714 (-8.79)	-0.406 (-3.17)	-0.135 (-1.08)	-0.482 (-6.89)	-0.487 (-2.95)	-0.136 (-0.84)	-0.628 (-6.95)	-0.359 (-2.54)	-0.102 (-0.72)	-0.459 (-5.81)
2	-0.060 (-0.70)	0.068 (0.78)	-0.228 (-4.69)	-0.109 (-1.37)	-0.029 (-0.35)	-0.143 (-3.1)	-0.197 (-1.96)	-0.084 (-0.80)	-0.203 (-3.48)	-0.166 (-1.79)	-0.079 (-0.81)	-0.156 (-2.89)
3	-0.064 (-1.07)	-0.069 (-1.09)	0.009 (0.24)	-0.075 (-1.27)	-0.076 (-1.23)	0.003 (0.08)	-0.069 (-0.96)	-0.050 (-0.66)	-0.035 (-0.82)	-0.062 (-0.86)	-0.063 (-0.83)	0.001 (0.02)
4	-0.001 (-0.02)	-0.117 (-1.85)	0.207 (5.85)	0.008 (0.13)	-0.075 (-1.25)	0.147 (4.41)	-0.007 (-0.09)	-0.080 (-1.07)	0.131 (3.11)	0.021 (0.31)	-0.038 (-0.54)	0.105 (2.67)
Winner	0.206 (2.22)	0.07 (0.75)	0.242 (4.59)	0.176 (2.37)	0.115 (1.48)	0.109 (2.52)	0.257 (2.48)	0.133 (1.25)	0.221 (3.70)	0.172 (1.90)	0.113 (1.19)	0.106 (1.99)
Winner-Loser	0.669 (3.03)	0.133 (0.64)	0.956 (8.17)	0.581 (3.33)	0.250 (1.44)	0.591 (6.11)	0.744 (3.08)	0.269 (1.13)	0.849 (6.36)	0.532 (2.60)	0.215 (1.04)	0.565 (4.87)

Table 2-7 Momentum Portfolios on NYSE/Amex/Nasdaq

This table reports the time-series regression of the price momentum portfolios. The sample includes all common stocks in NYSE/Amex/Nasdaq in the years 1984-2007 with valid CRSP data. The portfolios are formed based on the previous J months returns (skipping the last month) and held for K months. The dependent variables are the excess returns for momentum quintiles or the zero-investment portfolios returns with long positions on winners firms and short positions on losers firms. The value-weighted portfolios returns are regressed on the contemporaneous three Fama-French risk factors (MKT , SMB , and HML) and the information risk factor $ECINF$. MKT , the excess return on the market, is the value-weighted return on all NYSE/Amex/Nasdaq stocks minus the one-month Treasury bill rate. SMB (Small minus Big) is the return to size factor-mimicking portfolio. HML (High minus Low) is the return to book-to-market factor-mimicking portfolio. $ECINF$ is the return to information risk factor-mimicking portfolio. The table reports the alphas of the Fama-French three-factor model, the alphas and the $ECINF$ factor loadings of the Fama-French three-factor model augmented by $ECINF$. The t -statistics are in parentheses.

	J=6, K=6			J=6, K=12			J=12, K=6			J=12, K=12		
	Alpha FF3	Alpha ECINF	ECINF	Alpha FF3	Alpha ECINF	ECINF	Alpha FF3	Alpha ECINF	ECINF	Alpha FF3	Alpha ECINF	ECINF
Loser	-0.614 (-3.38)	-0.131 (-0.78)	-0.863 (-9.15)	-0.462 (-3.08)	-0.136 (-0.93)	-0.581 (-7.13)	-0.522 (-2.78)	-0.103 (-0.57)	-0.749 (-7.38)	-0.344 (-2.19)	-0.048 (-0.31)	-0.528 (-6.06)
2	-0.133 (-1.51)	0.029 (0.33)	-0.290 (-5.89)	-0.153 (-1.98)	-0.046 (-0.58)	-0.191 (-4.33)	-0.243 (-2.42)	-0.102 (-0.99)	-0.252 (-4.38)	-0.165 (-1.85)	-0.057 (-0.62)	-0.193 (-3.75)
3	-0.038 (-0.74)	-0.014 (-0.26)	-0.043 (-1.41)	-0.052 (-1.03)	-0.044 (-0.84)	-0.013 (-0.45)	-0.084 (-1.24)	-0.063 (-0.88)	-0.038 (-0.95)	-0.056 (-0.86)	-0.056 (-0.81)	0.000 (0.00)
4	0.027 (0.51)	-0.078 (-1.47)	0.187 (6.31)	0.024 (0.51)	-0.034 (-0.69)	0.104 (3.79)	0.022 (0.32)	-0.053 (-0.76)	0.134 (3.41)	0.046 (0.76)	-0.003 (-0.04)	0.088 (2.45)
Winner	0.374 (3.57)	0.224 (2.09)	0.269 (4.50)	0.264 (3.22)	0.221 (2.55)	0.078 (1.62)	0.361 (3.23)	0.25 (2.15)	0.198 (3.05)	0.241 (2.46)	0.208 (2.02)	0.058 (1.00)
Winner-Loser	0.989 (3.79)	0.354 (1.43)	1.132 (8.17)	0.726 (3.56)	0.356 (1.75)	0.660 (5.79)	0.884 (3.19)	0.353 (1.29)	0.947 (6.17)	0.585 (2.55)	0.257 (1.10)	0.586 (4.48)

Table 2-8 Post-Earnings-Announcement Drift on NYSE/Amex

This table reports the time-series regression of the earnings momentum portfolios. The earnings surprise (*SUE*) is defined as the actual earnings minus the expected earnings, scaled by stock price. We estimate *SUE* using both a rolling seasonal random walk (SRW) model and analysts' forecasts. *SUEs* for stock *i* in month *t* is defined as $(E_{i,t} - E_{i,t-4})/\sigma_{i,t}$, where $E_{i,t}$ is the quarterly earning most recently announced as of month *t* (not including announcements in month *t*) for firm *i*, and $\sigma_{i,t}$ is the standard deviation of earnings changes in the prior eight quarters. *SUE2* is similar to *SUE1*, except that the earnings in *SUE2* are exclusive of special items. *SUE3* is the earnings surprise based on I/B/E/S reported analysts' forecasts and actual earnings after replacing $E_{i,t-4}$ with a measure of analyst expectation. The sample includes common stocks in NYSE/Amex in years 1984-2007. The dependent variables are value-weighted excess returns for *SUE* quintiles or zero-investment portfolios returns with long positions on high *SUE* firms and short positions on low *SUE* firms. The portfolios returns are regressed on the contemporaneous three Fama-French risk factors (*MKT*, *SMB*, and *HML*) and the information risk factor *ECINF*. *MKT*, the excess return on the market, is the value-weight return on all NYSE/Amex/Nasdaq stocks minus the one-month Treasury bill rate. *SMB* (Small minus Big) is the return to size factor-mimicking portfolio. *HML* (High minus Low) is the return to book-to-market factor-mimicking portfolio. *ECINF* is the return to information risk factor-mimicking portfolio. The table reports the alphas of the Fama-French three-factor model, the alphas and the *ECIN* factor loadings of the Fama-French three-factor model augmented by *ECIN* factor. The *t*-statistics are in parentheses.

	SUE1			SUE2			SUE3		
	Alpha FF3	Alpha ECINF	ECINF	Alpha FF3	Alpha ECINF	ECINF	Alpha FF3	Alpha ECINF	ECINF
Low	-0.283 (-2.54)	-0.038 (-0.35)	-0.437 (-7.24)	-0.277 (-1.95)	0.029 (0.21)	-0.546 (-7.05)	-0.154 (-1.35)	-0.001 (-0.01)	-0.280 (-4.34)
2	-0.128 (-1.28)	0.016 (0.16)	-0.256 (-4.50)	-0.156 (-1.55)	-0.043 (-0.41)	-0.202 (-3.47)	-0.174 (-1.86)	-0.117 (-1.2)	-0.104 (-1.92)
3	-0.036 (-0.44)	-0.066 (-0.76)	0.054 (1.09)	-0.040 (-0.43)	-0.074 (-0.74)	0.060 (1.07)	-0.012 (-0.13)	0.017 (0.17)	-0.053 (-0.96)
4	0.001 (0.02)	-0.009 (-0.10)	0.018 (0.37)	-0.072 (-0.80)	-0.123 (-1.30)	0.090 (1.70)	-0.023 (-0.32)	0.037 (0.49)	-0.109 (-2.63)
High	0.146 (1.65)	0.064 (0.70)	0.146 (2.83)	0.128 (1.24)	0.032 (0.30)	0.170 (2.85)	0.210 (2.83)	0.196 (2.51)	0.025 (0.58)
High-Low	0.429 (2.83)	0.102 (0.69)	0.582 (7.06)	0.405 (2.27)	0.004 (0.02)	0.716 (7.44)	0.365 (2.78)	0.198 (1.48)	0.306 (4.11)

Table 2-9 Post-Earnings-Announcement Drift on NYSE/Amex/Nasdaq

This table reports the time-series regression of the earnings momentum portfolios. The earnings surprise (*SUE*) is defined as the actual earnings minus the expected earnings, scaled by stock price. We estimate *SUE* using both a rolling seasonal random walk (SRW) model and analysts' forecasts. *SUEs* for stock *i* in month *t* is defined as $(E_{i,t} - E_{i,t-4})/\sigma_{i,t}$, where $E_{i,t}$ is the quarterly earning most recently announced as of month *t* (not including announcements in month *t*) for firm *i*, and $\sigma_{i,t}$ is the standard deviation of earnings changes in the prior eight quarters. *SUE2* is similar to *SUE1*, except that the earnings in *SUE2* are exclusive of special items. *SUE3* is the earnings surprise based on I/B/E/S reported analysts' forecasts and actual earnings after replacing $E_{i,t-4}$ with a measure of analyst expectation. The sample includes common stocks in NYSE/Amex/Nasdaq in years 1984-2007. The dependent variables are value-weighted excess returns for *SUE* quintiles or zero-investment portfolios returns with long positions on high *SUE* firms and short positions on low *SUE* firms. The portfolios returns are regressed on the contemporaneous three Fama-French risk factors (*MKT*, *SMB*, and *HML*) and the information risk factor *ECINF*. *MKT*, the excess return on the market, is the value-weight return on all NYSE/Amex/Nasdaq stocks minus the one-month Treasury bill rate. *SMB* (Small minus Big) is the return to size factor-mimicking portfolio. *HML* (High minus Low) is the return to book-to-market factor-mimicking portfolio. *ECINF* is the return to information risk factor-mimicking portfolio. The table reports the alphas of the Fama-French three-factor model, the alphas and the *ECIN* factor loadings of the Fama-French three-factor model augmented by *ECIN* factor. The *t*-statistics are in parentheses.

	SUE1			SUE2			SUE3		
	Alpha FF3	Alpha ECINF	ECINF	Alpha FF3	Alpha ECINF	ECINF	Alpha FF3	Alpha ECINF	ECINF
Low	-0.321 (-2.4)	-0.032 (-0.25)	-0.516 (-7.09)	-0.309 (-2.09)	0.000 (0.00)	-0.552 (-6.8)	-0.182 (-1.56)	0.040 (0.35)	-0.406 (-6.37)
2	-0.095 (-0.91)	0.103 (1.00)	-0.354 (-6.12)	-0.169 (-1.6)	-0.01 (-0.09)	-0.284 (-4.72)	-0.205 (-2.27)	-0.167 (-1.77)	-0.069 (-1.3)
3	-0.026 (-0.38)	-0.041 (-0.56)	0.026 (0.64)	0.000 (0.00)	0.028 (0.29)	-0.049 (-0.91)	0.055 (0.6)	0.102 (1.06)	-0.087 (-1.63)
4	0.052 (0.72)	0.097 (1.28)	-0.08 (-1.88)	-0.074 (-0.95)	-0.126 (-1.54)	0.093 (2.03)	0.048 (0.56)	0.203 (2.42)	-0.283 (-6.07)
High	0.234 (3.38)	0.183 (2.52)	0.092 (2.28)	0.229 (2.57)	0.229 (2.43)	0.000 (0.00)	0.198 (2.57)	0.207 (2.56)	-0.017 (-0.38)
High-Low	0.556 (3.38)	0.215 (1.33)	0.608 (6.75)	0.538 (3.05)	0.229 (1.30)	0.552 (5.58)	0.379 (2.64)	0.167 (1.16)	0.388 (4.82)

Table 2-10 Seasoned Equity Offerings Portfolios

This table reports the *ECIN* characteristics and the time-series regressions of SEO portfolios returns on the three Fama and French (1993) factors (*MKT*, *SMB*, and *HML*) and the information risk factor *ECINF*. The sample includes seasoned equity offerings portfolios formed in years 1984-2007 on the common stocks in NYSE/Amex/Nasdaq. Portfolios returns are value-weighted. *MKT*, the excess return on the market, is the value-weight return on all NYSE/Amex/Nasdaq stocks minus the one-month Treasury bill rate. *SMB* (Small minus Big) is the return to size factor-mimicking portfolio. *HML* (High minus Low) is the return to book-to-market factor-mimicking portfolio. *ECINF* is the return to information risk factor-mimicking portfolio. Panel A reports the *ECIN* characteristics in years before and after SEO for NYSE and Amex SEO firms. Panel B reports the time-series regression for the SEO portfolios (the *t*-statistics are in parentheses.)

Panel A *ECIN* around SEO year

	<i>ECIN</i>	Diff(Post-Pre)	<i>t</i> (Diff)
Pre-SEO year	-0.205		
3-year Post-SEO	-0.260	-0.055	-11.39
5-Year Post-SEO	-0.277	-0.072	-15.54

Panel B Time-Series Regressions

	Intercept	<i>MKT</i>	<i>SMB</i>	<i>HML</i>	<i>ECINF</i>	Adj. R ²
Pre-SEO year	1.686 (11.61)	1.063 (28.3)	0.293 (6.27)	0.009 (0.16)		0.804
	1.607 (10.47)	1.094 (25.86)	0.328 (6.33)	0.015 (0.26)	0.132 (1.55)	0.805
3-year Post-SEO	-0.330 (-2.59)	1.193 (36.09)	0.128 (3.12)	0.058 (1.18)		0.857
	-0.199 (-1.49)	1.142 (31.04)	0.070 (1.56)	0.049 (1.00)	-0.221 (-2.96)	0.860
5-Year Post-SEO	-0.253 (-2.24)	1.174 (40.20)	0.139 (3.83)	0.115 (2.65)		0.878
	-0.144 (-1.22)	1.132 (34.77)	0.091 (2.28)	0.108 (2.50)	-0.182 (-2.76)	0.881

Table 2-11 Time Series Regressions of Share Issue Portfolios

This table reports the coefficients of Fama-French regressions of share issue portfolios. The regression is run on the sample period from January 1984 through December 2007 for firms in NYSE/Amex/Nasdaq with shares outstanding data in CRSP. *ISSUE* are the portfolios with long position on the firms of positive *ISSUE*_{-17,-6} including SEO and short positions on the firms of negative *ISSUE*_{-17,-6}, *NON-SEO* are the portfolios with long positions on the firms of positive *ISSUE*_{-17,-6} excluding SEO and short positions on the firms of negative *ISSUE*_{-17,-6}. Portfolios are formed in each month and held for 3 years. The value-weighted monthly returns are run on the Fama-French three-factor (*MKT*, *SMB*, and *HML*) model or the three-factor model augmented with *ECINF*. *MKT*, the excess return on the market, is the value-weight return on all NYSE, Amex, and Nasdaq stocks minus the one-month Treasury bill rate. *SMB* (Small minus Big) is the return on the small portfolios minus the return on the big portfolios. *HML* (High minus Low) is the return on the value portfolios minus the return on the growth portfolios. *ECINF* is the return on the high information risk portfolios minus the return on the low information risk portfolios. The sample period is from January 1984 to December 2007 (the *t*-statistics are in parentheses).

	Intercept	MKT	SMB	HML	ECINF	Adj. R ²
ISSUE	-0.205	0.108	0.182	-0.250		0.515
	(-2.33)	(4.79)	(6.55)	(-7.46)		
	-0.120	0.065	0.130	-0.263	-0.163	0.534
	(-1.34)	(2.59)	(4.16)	(-7.95)	(-3.47)	
NON-SEO	-0.183	0.100	0.172	-0.255		0.508
	(-2.10)	(4.49)	(6.25)	(-7.66)		
	-0.103	0.060	0.123	-0.267	-0.153	0.525
	(-1.16)	(2.41)	(3.97)	(-8.12)	(-3.28)

Figure 2-1 Time Series of SEO Listed by Exchange at the Time of Offering

The sample of seasoned equity offerings was identified in the SDC for the period 1979-2007. Offerings are listed by the exchange on which the firm was trading at the time of offering.

3. THE EFFECT OF REGULATION FAIR DISCLOSURE (REG FD) ON INFORMATION ASYMMETRY AND THE COST OF CAPITAL: REVISITED

3.1 Introduction

The U.S. Securities and Exchange Commission (SEC) approved Regulation Fair Disclosure (Reg FD) in October 2000. Aimed at reducing the information disparities between individual and institutional market participants, Reg FD requires that companies release material information to all investors simultaneously. By prohibiting selective disclosure, Reg FD should improve information flow, reduce information asymmetry and increase public confidence in the financial markets. However, opponents argue that companies will reduce the quality and quantity of the disclosure after Reg FD if the public disclosure is too costly or if firms are afraid of information being abused by rival companies. After Reg FD came into effect on October 23, 2000, an enormous amount of literature discusses to what extent the regulation has changed the information environment. As the purpose of Reg FD is to regulate the communication process between firms and financial analysts, many papers examine the change in the informativeness of analyst forecasts after the introduction of Reg FD. These papers (e.g. Heflin, Subramanyam and Zhang (2003), Bailey, Li, Mao and Zhong (2003), Mohanram and Sunder (2006), Gintschel and Markov (2004), Agrawal, Chadha and Chen (2006), and Irani and Karamanou (2003)) find that after the implementation of Reg FD, 1) analysts with preferential links to firms in the pre-FD period are unable to sustain the high-level forecasting accuracy, and that the increased

dispersion of analysts' forecasts suggests a greater disagreement among analysts when they have to rely more on "idiosyncratic information discovery", that 2) firms increase the voluntary disclosure without reducing the information quality of the disclosure, and that 3) stock prices are more informative and trading volume increases.

Besides financial analysts, another target of firms' selective disclosure is institutional investors. Ke, Petroni and Yu (2008) document that pre-FD transient institutional investors have abnormal selling of stocks preceding a bad news break¹⁵, while this trading behavior diminishes or disappears after Reg FD. As the institutional investors are typically informed traders, and the abnormal selling preceding bad news breaks are certainly informed trading, this evidence indicates that Reg FD reduces the possibility of informed trading. Thus these papers support that Reg FD has ultimately served its intended purpose to decrease (increase) private (public) information and consequently reduce information asymmetry. However, as these papers only focus on firms with at least two analysts following or on firms with large institutional holdings, the sample is biased toward the large firms.

Another related paper is Jorion, Liu and Shi (2005), who find that the credit rating changes are more informative in the post-FD period. Although Reg FD restricts nonpublic disclosure to financial analysts or investment professionals, it does not affect the disclosure to credit rating agencies. Therefore the information content of credit ratings increases after the implementation of Reg FD. In addition, their paper documents that the informational effects are more pronounced for firms with greater

¹⁵ "Bad new breaks are defined as breaks that are by growth firms, preceded by longer strings of consecutive earnings increases, followed by longer strings of consecutive earnings decreases, and associated with larger declines in earnings." (Ke, Petroni and Yu (2008))

analyst following and for larger firms, which are more likely to practice selective disclosure in the pre-FD period. Another two papers provide direct evidence on the reduced information asymmetry in the NYSE firms. Eleswarapu, Thompson and Venkataraman (2004) employ the percentage effective spread and the percentage price impact as the proxies of trading costs to suggest that information asymmetry declines after Reg FD for the NYSE firms on trading days surrounding the release of earnings information. Similarly, Collver (2007) documents a significant decrease in Hasbrouck (1991b) summary informativeness statistics after Reg FD for a sample of large NYSE firms (S&P 100 components). In sum, a slew of papers in the literature have shown that Reg FD increases firms' public voluntary disclosure and decreases information asymmetry, although the samples in those papers are biased towards large firms.

However, when the sample is extended to the Nasdaq firms or smaller firms, the results are different. With the sample of NYSE and Nasdaq firms that have analysts forecasts data in the First Call Historical Database, Gomes, Gorton and Madureira (2007) find that the adoption of Reg FD causes a significant "reallocation of information-producing resources", resulting in welfare loss for small firms. In particular, small firms lose 17 percent of their analyst followings, while big firms gain 7 percent. Big firms' voluntarily disclosing information through pre-announcements increases 100% after Reg FD, while small firms do not increase the pre-announcements significantly. Small firms experience higher forecast errors and volatility at earnings announcements, while large firms experience no significant increases. With a dummy indicating the post-FD period in the Fama-French three-factor model, they find a significant increase in the cost of capital for the small and

mid-size firms. This evidence is consistent with Bushee, Matsumoto and Miller (2004), who study whether firms that host closed conference calls in the pre-FD period continue to host (open) conference calls in the post-FD period. Their results reveal that although the percentage of firms that discontinue hosting conference calls is low, such firms tend to be smaller firms with low institutional ownership. They argue that for small firms, the costs of hosting a public call are large while the benefits of hosting calls are small.

Although previous literature provides detailed analyses on the changes in the information environment of firms after the implementation of Reg FD, few studies have directly measured the economic impact of the regulation by examining how these changes in the information environment affect the cost of capital. Duarte, Han, Harford and Young (2008) (DHHY thereafter) is an exception. They analyze the impact of Reg FD on information asymmetry and the cost of capital, based on the information asymmetry measure *PIN* developed by Easley, Kiefer, O'Hara and Paperman (1996). They find that NYSE/Amex firms have little/no changes in information asymmetry but that Nasdaq firms have significantly increases in information asymmetry and related increases in the cost of capital by 0.17% after Reg FD. However, the lack of improvement in information environment for NYSE/Amex firms found in DHHY seems at odds with the improvement cited in earlier studies (cf. Eleswarapu, Thompson and Venkataraman (2004), Gomes, Gorton and Madureira (2007) and Collver (2007).) The source of this discrepancy may be linked to the fact that *PIN* is not an effective measure of information asymmetry or risk. Duarte and Young (2009) argue that *PIN* is priced because it measures the illiquidity instead of information risk and that the pricing effect of *PIN* is negligible when Amihud (2002) illiquidity measure is controlled. Mohanram and Rajgopal (2008) find the pricing

effect of *PIN* disappears in the sample years of 1989-2002. The disappearance of the pricing effect of *PIN* in the more recent samples may be caused by the explosive increase in trading frequency we have seen in recent years. The noisy estimation of *PIN* is mainly caused by the convergence problem in the maximum likelihood estimation for highly frequently traded stocks. Easley, Hvidkjaer and O'Hara (2005) document that this problem is much more severe in large firms as their stocks are traded more frequently. Consequently, the findings of DHHY are not conclusive, especially for NYSE/Amex firms.

In this paper, we revisit DHHY with a more consistent and precise information asymmetry measure *ECIN* based on the price discovery of large trades. As shown in the first essay, the pricing effect of *ECIN* is robust to different time periods and to controlling for various measures of illiquidity, and subsumes the pricing effect of *PIN*. Unlike DHHY, we find that Reg FD significantly reduces information asymmetry and consequently reduces the cost of capital by 0.15% for NYSE/Amex firms. This result is closely related with Eleswarapu, Thompson and Venkataraman (2004), Gomes, Gorton and Madureira (2007) and Collver (2007) who find a significant decrease in the information asymmetry for the NYSE/Amex firms. On the other hand, we find similar results with DHHY for Nasdaq firms that Reg FD increases information asymmetry and the cost of capital, which is also consistent with Gomes, Gorton and Madureira (2007), who find that small firms face a higher cost of capital after the adoption of Reg FD. A further analysis reveals that the different effects in NYSE/Amex and Nasdaq firms attribute to the different characteristics. We pool all firms together and run a cross-sectional regression of change in the cost of capital on a

dummy variable equal one for Nasdaq firms. We find significant negative coefficient on the dummy, which is consistent with the subsample tests. But the dummy variable is not significant when we control firm size, book-to-market and *ECIN*. Moreover, the change in the cost of capital has significantly negative coefficients on these three firm characteristics. Thus, Reg FD successfully reduces information asymmetry and the cost of capital for large firms. However, as public disclosure, e.g. a conference call open to all investors, is too costly or not so beneficial for small firms, Reg FD creates “the chilling effect” of information flow for these firms. On the other hand, value firms or firms with high *ECIN* typically have higher information risk, so they benefit more from the regulation.

The contribution of this paper goes beyond the debate on the effectiveness and the associated costs and benefits of Reg FD. This study provides confirming evidence of the superiority of *ECIN* over *PIN* as the information risk measure. In essay one, we have shown that *ECIN* subsumes *PIN* in the pricing impact. Reg FD provides an excellent opportunity to compare the efficacy of *ECIN* and *PIN* as the information risk measure. First, the main purpose of Reg FD is to “level the playing field” by transforming private information to public information. Therefore this event has an obvious impact on information environment. Secondly, this event has a determinate effective date and affects all firms simultaneously, therefore facilitating the cross-sectional comparison of firms with different characteristics. Thirdly, a large body of literature documenting the significant changes in information environment around the implementation of Reg FD varies across firms of different characteristics and exchanges. An effective information risk measure should pass the simple test that it exhibits the corresponding changes to the changes in information environment as documented in the literature. Based on the studies we cite earlier, it is clear that while

both *ECIN* and *PIN* pass this test in the sample of Nasdaq firms, only *ECIN* yields the results that are consistent with the literature in the NYSE/Amex sample. As previously discussed, *PIN* is especially noisy for the actively traded stocks which are more likely listed on NYSE/Amex. The difference not only explains why the pricing impact of *PIN* is subsumed by that of *ECIN* in the NYSE/Amex sample, but also suggest that we should be cautious about, if not re-examine, the results in those studies that have relied on *PIN* as the information risk measure, especially when the samples consist mostly of NYSE/Amex firms.

The remainder of the paper is organized as follows. Section 2 describes the proxies for information asymmetry and dissemination. Section 3 describes the data and sample. Section 4 introduces the asset pricing tests. Section 5 tests the changes in information asymmetry and the cost of capital around Reg FD. Section 6 concludes.

3.2 Proxy for Information Asymmetry and Proxies for Information Dissemination

The proxy for information asymmetry is the information risk measure *ECIN* developed in the first essay. For each NYSE/Amex/Nasdaq stock with valid trading data in ISSM or NYSE TAQ, we apply the vector error-correction model of equation (1.1) in the first essay to estimate the information risk measure *ECIN* based on the price discovery of large trades. As the firm size and the trading size of Nasdaq stocks can be largely different from those of NYSE/Amex stocks, we classify the large and small trades separately for NYSE/Amex and for Nasdaq firms.

Following DHHY, we create the proxies for information dissemination as *CDUM*, *SDUM* and *ADUM*. The first proxy *CDUM* is a dummy equal to one if a firm has a credit rating. Because the analyses of rating agencies convey information about the firm to the market, a firm with a credit rating would have greater information dissemination. The second proxy *SDUM* is a dummy equal to one if the firm's institutional ownership percentage is higher than the full sample median ownership percentage. As institutional investors are more likely to be informed traders, large fraction of institutional investors indicates a greater dissemination of information. The third proxy *ADUM* is a dummy equal to one if the firm is covered by analysts. Intuitively, analysts dispense and spread the information to investors and therefore increase the dissemination of information.

3.3 Data and Sample

For the asset pricing test, we include all ordinary common stocks listed on NYSE/Amex/Nasdaq with valid trading data in ISSM or NYSE TAQ databases. *ECIN* is estimated yearly from 1983 to 2005 for NYSE/Amex firms and from 1987 to 2005 for Nasdaq firms. We also require valid stock returns and market equity data in CRSP monthly files and book equity data in Compustat annual files. We exclude stocks with less than six months of records in CRSP or stocks with negative or zero shares outstanding. We also delete REIT, stocks of companies incorporated outside U.S. and closed-end funds. Beta (β) is estimated as the post-ranking portfolio betas of Fama and French (1992) in the previous calendar year based on 40 portfolios. Size is the logarithm of the market value of equity at the previous year end, calculated as the product of price and shares outstanding. Book-to-market (*BM*) is logarithm of the

book value divided by the market value at the end of the previous year and winsorized at 0.5% and 99.5%. In addition, we delete firms with negative book values. Amihud (2002)'s illiquidity measure is calculated as

$$ILLIQ_{iy} = \left(\frac{1}{D_{iy}}\right) \sum_{i=1}^{D_{iy}} |R_{iyd}| / VOLD_{iyd} \quad (3.1)$$

where D_{iy} is the number of days for which data are available for stock i in year y ; R_{iyd} is the return on stock i on day d of year y ; $VOLD_{iyd}$ is the corresponding daily volume in dollars.

For the *ECIN* estimation, we exclude ISSM/TAQ intraday data with non-positive price, for trades outside regular trading hours or with irregular terms.¹⁶ Open and close trades are also excluded since those trades are aggregated artificially as large size trades and do not necessarily indicate any private information. To exclude possible reporting errors, we also delete the trade with price change of more than 50% from the previous trade and with reverse price change of more than 50% in the immediately following trade. We merge ISSM with CRSP by ticker symbols, and TAQ with CRSP by CUSIPs. According to Sadka (2006), this merging method yields the highest matching rate. Moreover, Easley, Kiefer, O'Hara and Paperman (1996) argue that the large trade tends to have multiple participants on one side of the trade.

¹⁶ We delete trades and quotes outside the regular trading time 9:30-16:00. For ISSM data, we only include trades and quotes with their condition code as blank or “*”. For TAQ quote data, we exclude quotes with the mode as 4,7,9,11,13,14,15,19,20,27, or 28. For TAQ trade data, we only include trades with the correction indicator as 0 or 1 and the condition code as blank or “*”.

Such trade is typically reported as multiple transactions in reporting conventions, while actually only one trade happens. To eliminate this problem, we aggregate all trades occurring within 5 seconds of each other at the same price and with no intervening quote revisions.

For the information dissemination proxies, we collect the credit rating data from the Compustat industrial quarterly files, which are available from year 1984. *CDUM* takes the value of one if the firm has an S&P Long-Term Domestic Issuer Credit Rating and/or an S&P Short-Term Issuer Credit Rating and/or an S&P Subordinated Debt Rating. Institutional ownership data is from Thomson-Reuters Institutional holdings, which gathers information from institutional 13F SEC filings. The filings are submitted quarterly and include institutional holdings in every U.S. firm, as long as the holdings are more than \$200,000 or 10,000 shares. In each quarter, we aggregate the institutional holdings for each firm. If there is more than one record of the same institution manager holding one firm in the same quarter, we only keep the record with the latest file date. For the sample of firms with institutional holding data, the fraction of institutional holding is winsorized at the 95th percentile to eliminate recording errors. Institutional holdings are set to zero for firms without matched 13F data. We obtain data on analyst coverage from the I/B/E/S Historical Summary files. For each stock i in any given quarter t , *ADUM* takes the value of one if at least one analyst provide quarter-1- to quarter-4-ahead (*FPI* has the value of “6”, “7”, “8” or “9”) earnings estimation. If no I/B/E/S value is available, *ADUM* takes the value of zero. As the *ADUMs* are largely zero in the year of 1983, the asset pricing test for NYSE/Amex firms with *ADUM* starts from 1984.

Table 3-1 reports the sample statistics for the variables. Nasdaq firms have larger *ECIN* (closer to zero), indicating higher information asymmetry, than do NYSE/Amex firms. Given that Nasdaq firms are smaller than NYSE/Amex, this result is consistent with the first essay's findings that *ECIN* is negatively correlated with firm size. Table 3-2 reports the correlations between the information dissemination proxies and *ECIN*. The three information dissemination proxies are positively correlated with each other, and they are negatively correlated with *ECIN*.

3.4 Asset Pricing Tests

Following DHHY, we estimate the asset pricing model including *ECIN* and the interactions between *ECIN* and the proxies for information dissemination. Every month, we estimate the following cross-sectional regression with the *ECIN* measure

$$r_i = a + (\theta_1 + \theta_2 ID_i) ECIN_i + \gamma_1 \beta_i + \gamma_2 SIZE_i + \gamma_3 BM_i + \varepsilon_i \quad (3.2)$$

where $ID_i \in \{0, ADUM_i, SDUM_i, CDUM_i\}$.

According to Easley and O'Hara (2004), the stock return is positively correlated with information asymmetry since privately informed trading increases information risk. On the other hand, higher information dissemination reduces information asymmetry and hence the compensation for information risk. Thus we expect that θ_1 is positive and that θ_2 is negative. Like DHHY and Easley, Hvidkjaer and O'Hara (2002), we estimate the coefficients in equation (3.2) with the weighted

least-squares procedure of Litzenberger and Ramaswamy (1979). Panel A of Table 3-3 reports the coefficients for NYSE/Amex stocks and Panel B reports the coefficients for Nasdaq stocks. The coefficient on *ECIN* for Nasdaq stocks is 0.889 (t -statistics=2.50), which is higher than the coefficient on *ECIN* for NYSE/Amex stocks at 0.444 (t -statistics=3.18), suggesting that Nasdaq stocks are compensated more for information asymmetry risk. Similarly, *DHHY* shows a higher coefficient on *PIN* for Nasdaq stocks than that for NYSE/Amex stocks. The coefficients on the interactions of information disseminations and *ECIN* are mostly negative. The coefficient on *ECIN* remains positive and significant in the regressions including the interactions.

Duarte and Young (2009) argue that the pricing effect of *PIN* is subsumed by the Amihud (2002) illiquidity measure. We include the Amihud (2002) illiquidity measure into the cross-sectional regression and report the weighted least-squares estimation in Table 3-4. The results in Panel A for NYSE/Amex firms are similar to that in the first essay; the coefficients on the illiquidity measure are insignificant and slightly negative while the coefficients on *ECIN* remain significantly positive. Panel B reports the results for Nasdaq firms where the coefficients on *ECIN* remain almost unchanged from the regressions reported in Table 3-3 and the coefficients on the illiquidity measure are insignificant. The significance of *ECIN* on Nasdaq firms with controlling for the illiquidity measure is essential to illustrate the pricing effect of information risk, given that Nasdaq-listed stocks are typically small and illiquid.

As shown in Table 3-2, *ECIN* have significantly negative correlations with the information dissemination proxies. In addition, the information dissemination proxies can also be the weak proxies for information asymmetry. Analyst coverage, for example, are negatively related to adverse selection costs (Brennan and

Subrahmanyam (1995).) Therefore one may argue that the interaction effect between dissemination proxies and *ECIN* is spurious due to the fact that dissemination measures are simply noisy proxies for *ECIN*. Following DHHY, we include the squared value of *ECIN* ($ECIN^2$) in the cross-section regression to test whether the coefficients θ_2 on the interactions are simply picking up the spurious effect. Table 3-5 reports the results of the new regressions. We can observe that there are negligible changes in the coefficients for the interactions of *ECIN* and the information dissemination proxies. Therefore, we disprove the argument that *ADUM*, *CDUM* or *SDUM* are simply noisy proxies for *ECIN*.

3.5 Changes in Information Asymmetry and in the Cost of Capital around Regulation Fair Disclosure

We estimate the changes in *ECIN* with the modified VECM. We create a dummy which takes the value of one on the days post-FD and zero on the days pre-FD, and fit it to the VECM time-series regression. Including the dummy variable into the model of equation (1.1) gives,

$$\Delta P_{i,t}^L = c_i^L + \alpha_i^L Z_{i,t-1} + \alpha_i^{L,D} (POS \times Z_{i,t-1}) + \sum_{h=1}^q \phi_{i,h} \Delta P_{i,t-h}^L + \sum_{h=1}^q \delta_{i,h} \Delta P_{i,t-h}^S + \eta_{i,t}^L$$

$$\Delta P_{i,t}^S = c_i^S + \alpha_i^S Z_{i,t-1} + \alpha_i^{S,D} (POS \times Z_{i,t-1}) + \sum_{k=1}^q \lambda_{i,k} \Delta P_{i,t-k}^L + \sum_{k=1}^q \gamma_{i,k} \Delta P_{i,t-k}^S + \eta_{i,t}^S$$

(3.3)

where POS is a dummy variable which takes the value of one on the days post Reg FD and zero on the days before Reg FD. Therefore, $\alpha_i^{L,D}$ measures the increase in the price discovery of large trades, and hence the increase in information asymmetry, after Reg FD for stock i . We run the VECM in the period from October 23, 1999 to January 23, 2001, in which the post-FD period is from October 23, 2000 to January 23, 2001. Assuming the level of information asymmetry for a given company is stable over a period of one year, the longer estimation period before FD allows us more observations that enable us to get a more accurate estimation of $ECIN$. As the robustness check, we also run the regression with the shorter pre-Reg FD period, July 21, 2000 to January 23, 2001, as defined in DHHY. This estimation is based on fewer observations, but the magnitude and significance of the changes in $ECIN$ and in the cost of capital are quite similar to those reported in the paper. Following DHHY, the post-event period starts from the active day of Reg FD and ends when the decimalization comes into effect for NYSE/Amex. Because financial literature reveals that decimalization has significant effect on the information environment (e.g. Bacidore (1997, 2001), Chakravarty, Van Ness and Van Ness(2005), Chou and Chung(2004), Chung and Zhao (2005), and Gibson, Singh and Yerramilli (2003)), the sample period in this study ends before the decimalization to isolate the effect of decimalization and focus on the impact of Reg FD. Analyst following, institutional ownership, and credit rating are quarterly data. Following DHHY, we use the data for the calendar quarter from July to September 2000 as the pre-event values for $ADUM$, $SDUM$, and $CDUM$ and the data for the calendar quarter from October to December

2000 as the post-event values. Table 3-6 reports the average changes in *ECIN* and in the proxies for information dissemination. Following DHHY, we take the average separately for NYSE/Amex and Nasdaq. We find that *ECIN* significantly decreased after the adoption of Reg FD for NYSE/Amex stocks, which contradicts DHHY, who find insignificant changes in *PIN*. But similar to DHHY, we find significant increases in *ECIN* for the Nasdaq firms.

For each firm i , we calculate the change in the cost of capital resulting from the change in the firms' information environment before and after Reg FD as following

$$\Delta r_i = \theta_1 [ECIN_{After,i} - ECIN_{Before,i}] + \theta_2 [ECIN_{After,i} \times ID_{After,i} - ECIN_{Before,i} \times ID_{Before,i}] \quad (3.4)$$

where $ID_i \in \{0, ADUM_i, SDUM_i, CDUM_i\}$ and θ_1 and θ_2 are estimated in equation (3.2). We report the changes in the cost of capital calculated with θ_1 and θ_2 from the estimations in Table 3-3 without controlling for the effect of illiquidity or $ECIN^2$. The changes in the cost of capital are essentially identical if we instead use the parameters in Table 3-4 or Table 3-5 and include any significant effect of these two variables. In addition, to control for the effects of other events coincident with Reg FD, we run the cross-sectional regression:

$$\Delta r_i = c_0 + c_1 SIZE_i + c_2 BM_i + c_3 ECIN_i + c_4 \Delta VOL_i + c_5 \Delta LARGE_i + c_6 HITECH_i + e_i \quad (3.5)$$

Similar to DHHY, we include *SIZE*, *BM*, and *ECIN* in the regression to understand the cross-sectional determinants of Reg FD's effect on the cost of capital. *SIZE* is the logarithm of the market value of equity at the end of year 1999. *BM* is the logarithm of the ratio of the book equity value to the market value of equity at the end of year 1999. *ECIN* is estimated with the vector error-correction model (VECM) over the year of 1999. In addition, the contemporaneous collapse of the Internet bubble would have effects on information-based trading. We include the dummy variable *HITECH_i* which takes the value of one if the firm *i* is in the high-tech industry, according to the ten Fama and French industry classifications. ΔVOL_i and $\Delta LARGE_i$ captures the changes in trading volume that are impacted by market-wide abrupt price declines in the Internet-related stocks. Specifically, ΔVOL_i is the change in average monthly share volume, from the period of three months before Reg FD (July to September 2000) to the period of three months after Reg FD (October to December 2000), expressed in units of ten million shares. $\Delta LARGE_i$ is the change in the firms' monthly average number of trades with ten thousand shares or more, from the period three months before Reg FD to the period three months after Reg FD, expressed in units of one hundred trades. Except for the dummy variable, all the explanatory variables are demeaned by the mean of the respect variables in the sample of study. Therefore, the constant c_0 reflects the changes in the cost of capital for firms with average *SIZE*, *BM*, and *ECIN*, purged of effects other than Reg FD.

We estimate the changes in the cost of capital with and without interaction from *ECIN* and the information dissemination proxies. Table 3-7 reports the annualized change in the cost of capital for NYSE/Amex and Nasdaq subsamples with different information dissemination proxies. We report the average unadjusted (Δr in equation (3.4)) and the characteristic-adjusted changes (c_0 in equation (3.5)) in the

cost of capital. From Panel A, we can see the cost of capital significantly decreases for the NYSE/Amex firms. The change ranges from (-0.15%) to (-0.334%) and is highly significant. Panel B of Table 3-7 provides the results for Nasdaq firms. Similar to DHHY, the Nasdaq firms on average have an increased cost of capital. In addition, we find the changes in the cost of capital have a very significant and negative relation with firm size for Nasdaq. Therefore the larger (smaller) firms tend to reduce (increase) more or increase (reduce) less the cost of capital after the implementation of Reg FD. DHHY find a similar negative correlation between firm size and the changes in cost of capital for Nasdaq firms.

In addition, we find a significantly negative correlation for Nasdaq firms between the previous year *ECIN* and changes in the cost of capital: firms with high *ECIN* before Reg FD tend to experience less increase (maybe even reduction) in the cost of capital, while firms with low *ECIN* before Reg FD tend to experience more increase in the cost of capital. This is intuitive as Reg FD targets to reduce information asymmetry for the firms with high levels of privately informed trading by prohibiting selective disclosure to potential informed traders. However, firms that originally have low levels of information asymmetry enjoy fewer benefits from the implementation of Reg FD and still suffer extra costs from the required public disclosure or from stopping the selective disclosure that is often crucial for information dissemination before Reg FD. Similarly, DHHY find a significant negative correlation between *PIN* and the changes in the cost of capital for Nasdaq firms. Both DHHY and this paper do not find any significant correlation between the original levels of information asymmetry and the changes in the cost of capital in the

NYSE/Amex sample. Table 3-7 also reveals that the changes in the cost of capital for Nasdaq firms are more sensitive to controlling for the volume changes and the high-tech dummy, as most of the Internet-related firms are listed on Nasdaq and are more affected by the dot-com bubble. However, even after we purged all these market effects, the changes in the cost of capital related to Reg FD for Nasdaq firms are still significant.

The noteworthy difference between NYSE/Amex and Nasdaq firms and the difference between this paper and DHHY motivate us to investigate deeper into the sample. We pool NYSE/Amex and Nasdaq stocks together and define a dummy variable equal to one if a stock is listed in Nasdaq and zero otherwise. We regress the change in the cost of capital on the Nasdaq dummy and report the coefficients in Table 3-8. Without controlling for firm characteristics, there are significant positive coefficients on the Nasdaq dummy and negative coefficients on constant variable, consistent with the previous results. Then we control the firm size, book-to-market ratio and *ECIN*, which are demeaned by the mean of the respect coefficients of the whole sample. We find three out of four measures of changes in cost of capital give insignificant coefficients on both constant variable and dummy variable. This result suggests that the different changes in the cost of capital for NYSE/Amex and for Nasdaq stocks are due to the different firm characteristics. The impact of Reg FD is negatively related to firm size, book-to-market and *ECIN*. For the average size, book-to-market and *ECIN* firm, Reg FD does not significantly affect firm's cost of capital, no matter which exchange the firm is listed. But because Reg FD requires firms to disclosure information to all investors simultaneously, it reduces information asymmetry for the large firms which can afford the costly public disclosure, but it may also cause "the chilling effect" for the small firms which choose to stop any form of

disclosure after Reg FD. High book-to-market (value firm) and high *ECIN* firms are potentially subject to higher information risk. The aim of Reg FD is to “level the playing field” or to reduce information asymmetry, so it is not surprising that firms with higher information risk reduce the cost of capital more.

As the impact of Reg FD in the cost of capital permanently decreases (increases) the discount rates for the average NYSE/Amex (Nasdaq) firms, the price effects of the change is potentially large. Following DHHY, we employ the Vuolteenaho (2002) method to calculate the effect of the change in the discount rate on the percentage price change. The calculation is based on the Campbell and Shiller decomposition to represent the log of a stock’s price as

$$p_t \approx \sum_{l=0}^{\infty} \rho^l \Delta E[d_{t+l}] - \sum_{l=0}^{\infty} \rho^l \Delta E[r_{t+l}] \quad (3.6)$$

where ρ is the discount factor emerging from the linearization; d_{t+l} is the dividend at time $t+l$; r_{t+l} is the stock’s return at time $t+l$. Following DHHY, the parameter ρ takes the value of 0.967. Assuming Reg FD has no effect on future dividends, the percentage price increases from the decreases in the future expected discount rates of the average NYSE/Amex stock is at least

$$-\sum_{l=0}^{\infty} \rho^l \Delta E[r_{t+l}] = -\sum_{l=0}^{\infty} (0.967)^l (-0.0015) \approx 3.5\%$$

The percentage price decreases for Nasdaq stocks is at least

$$-\sum_{l=0}^{\infty} \rho^l \Delta E[r_{t+l}] = -\sum_{l=0}^{\infty} (0.967)^l 0.0018 \approx -4.2\%$$

Thus, if Reg FD is a complete surprise to the market in the month of implementation, it will at least cause an unexpected return for the average NYSE/Amex (Nasdaq) firm at 3.5% (-4.2%).

3.6 Conclusion

Based on the new information risk measure we developed in the first essay, *ECIN*, we reexamine the impact of Reg FD on firms' costs of capital through its impact on information risk. We find that *ECIN*, and the related cost of capital, has significantly decreased for the NYSE/Amex firms after the implementation of Reg FD. Nasdaq firms, on the other hand, have experienced increases in information asymmetry, and hence increases in the cost of capital after Reg FD. A further analysis reveals that the different effect for exchanges are because of the different characteristics of listed firms. In particular, large, high book-to-market and high *ECIN* firms benefit more from the implement of Reg FD.

Previous studies of this nature, like DHHY, that use *PIN* as the information risk measure are vulnerable to criticism. First, Duarte and Young (2009) argue that *PIN* actually captures illiquidity rather than information risk. Secondly, DHHY's result that there are no significant changes in information asymmetry for NYSE/Amex firms is in contradiction to a large amount of literature showing that Reg FD effectively reduces information asymmetry, especially for the relatively large firms. In contrast, this paper documents a significant reduction in *ECIN* that is consistent with the literature. This sharp contrast reaffirms our findings in the first essay that *ECIN* is a better information (asymmetry) risk measure than *PIN* in the NYSE/Amex sample,

which can be attributed to the fact that *PIN* estimation is noisy in the sample of large firms with high trading frequency.

DHHY's results that Reg FD increases the cost of capital for Nasdaq firms while leaving that of NYSE/Amex firms unchanged, would have lead one to conclude that Reg FD is a piece of harmful legislation as far as information risk and the cost of capital are concerned. Thus, our results are important as we show that while there is an increase in the cost of capital of Nasdaq firms, this increase is more than offset by the decrease in the cost of capital of NYSE/Amex firms. Our estimates show that Reg FD increases the market value of NYSE/Amex firms by 3.5%, while it decreases the market value of Nasdaq firms by 4.2%. Considering that NYSE/Amex stocks have much larger market capitalization than Nasdaq firms have, our results suggest that Reg FD is a piece of good legislation: not only does it level the playing field for investors as has been shown in the literature, but it also increases the wealth of the overall market through the reduction of the cost of capital.

Table 3-1 Summary Statistics

This table reports the mean, median, standard deviation, 25th percentile and 75th percentile for the variables in asset pricing tests. *ECIN* is the information risk measure estimated yearly following the model of equation (1.1). *SIZE* is the logarithm of the year-end market value of equity (in million dollars). *BM* is the logarithm of the book value divided by the market value. *BETA* is the post-ranking portfolio betas of Fama-French (1992) estimated using 40 portfolios. *ILLIQ* is the Amihud (2002) illiquidity measure. *ADUM* is equal to one if the firm has one or more analysts, and zero otherwise. *CDUM* is equal to one if the firm has a credit rating, and zero otherwise. *SDUM* is equal to one if the fraction of shares held by large institutions exceeds the median level for the entire sample, and zero otherwise. The NYSE and Amex sample includes all common stocks in years 1983-2006, except for *ADUM* and *CDUM*, which start from 1984. The Nasdaq sample includes all common stocks in the years 1987-2006. Panel A reports the summary statistics for NYSE/Amex and Panel B reports the summary statistics for Nasdaq.

Panel A NYSE/Amex

	Mean	Median	Std. Dev.	P25	P75
ECIN	-0.246	-0.226	0.212	-0.320	-0.097
SIZE	6.078	6.116	1.930	4.926	7.650
BM	-0.55	-0.507	0.718	-0.984	-0.150
BETA	1.193	1.154	0.291	0.954	1.340
ILLIQ(10 ⁶)	1.819	0.037	1.003	0.002	0.142
ADUM	0.722	1	0.448	0	1
CDUM	0.440	0	0.496	0	1
SDUM	0.597	1	0.491	0	1

Panel B Nasdaq

	Mean	Median	Std. Dev.	P25	P75
ECIN	-0.199	-0.123	0.200	-0.259	-0.072
SIZE	5.075	4.990	1.545	3.983	6.050
BM	-0.724	-0.651	0.810	-1.183	-0.194
BETA	1.272	1.216	0.355	0.997	1.563
ILLIQ(10 ⁶)	1.210	0.124	3.750	0.015	0.953
ADUM	0.659	1	0.474	0	1
CDUM	0.114	0	0.318	0	0
SDUM	0.395	0	0.489	0	1

Table 3-2 Correlations between *ECIN* and Information Dissemination Proxies

This table reports the correlation coefficients between the analyst dummy (*ADUM*), the share dummy (*SDUM*), the credit rating dummy (*CDUM*) and information asymmetry (*ECIN*), estimated yearly using the model of equation (1.1). The correlation coefficients are calculated over the period 1984 to 2005 for the sample of NYSE and Amex, and from 1987 to 2005 for the sample of Nasdaq. *ADUM* is equal to one if the firm has one or more analysts, and zero otherwise. *CDUM* is equal to one if the firm has a credit rating, and zero otherwise. *SDUM* is equal to one if the fraction of shares held by large institutions exceeds the median level for the entire sample, and zero otherwise. Panel A reports the correlation coefficients of NYSE/Amex and Panel B reports the correlation coefficients for Nasdaq. The *p*-values are in parentheses below the correlation coefficients.

Panel A NYSE/Amex

	ADUM	CDUM	SDUM
ECIN	-0.391 <.0001	-0.389 <.0001	-0.247 <.0001
ADUM		0.423 <.0001	0.359 <.0001
CDUM			0.214

Panel B Nasdaq

	ADUM	CDUM	SDUM
ECIN	-0.377 <.0001	-0.165 <.0001	-0.301 <.0001
ADUM		0.163 <.0001	0.384 <.0001
CDUM			0.173 <.0001

Table 3-3 Fama and MacBeth Regression Results

This table reports the time-series averages of the monthly estimated coefficients in cross-sectional regressions for NYSE/Amex firms and for Nasdaq firms, using the weighted least-square methodology suggested by Litzenberger and Ramaswamy (1979). The dependent variable is the percentage monthly return. *ECIN* is the information risk measure estimated yearly following the model of equation (1.1). *SIZE* is the logarithm of the year-end market value of equity (in million dollars). *BM* is the logarithm of the book value divided by the market value. *BETA* is the post-ranking portfolio betas of Fama-French (1992) estimated using 40 portfolios. *ADUM* is equal to one if the firm has one or more analysts, and zero otherwise. *CDUM* is equal to one if the firm has a credit rating, and zero otherwise. *SDUM* is equal to one if the fraction of shares held by large institutions exceeds the median level for the entire sample, and zero otherwise. Panel A reports the Fama-MacBeth regressions for NYSE/Amex common stocks in years 1984-2007, where the regressions involving *ADUM* and *CDUM* start from 1985. Panel B reports the Fama-MacBeth regressions for Nasdaq common stocks in years 1988-2006. The *t*-statistics are in the parentheses.

Panel A NYSE/Amex

	ECIN	Beta	SIZE	BM	ECIN×ADUM	ECIN×CDUM	ECIN×SDUM
(1)	0.444 (3.30)	-0.425 (-1.70)	0.151 (1.86)	0.112 (3.42)			
(2)	1.030 (2.04)	-0.455 (-1.81)	0.142 (1.76)	0.097 (3.00)	-0.052 (-0.11)		
(3)	0.885 (4.41)	-0.433 (-1.73)	0.136 (1.70)	0.101 (3.00)		-0.461 (-2.36)	
(4)	0.854 (4.54)	-0.468 (-1.89)	0.152 (1.87)	0.103 (3.19)			-0.403 (-2.27)

Panel B Nasdaq

	ECIN	Beta	SIZE	BM	ECIN×ADUM	ECIN×CDUM	ECIN×SDUM
(1)	0.889 (2.50)	-0.398 (-1.22)	0.380 (3.05)	0.149 (2.74)			
(2)	0.994 (2.41)	-0.41 (-1.26)	0.373 (2.98)	0.141 (2.36)	-0.366 (-0.67)		
(3)	0.871 (2.47)	-0.4 (-1.23)	0.384 (3.14)	0.152 (2.74)		-0.101 (-0.36)	
(4)	0.998 (2.52)	-0.409 (-1.27)	0.375 (3.01)	0.13 (2.33)			-0.233 (-0.83)

Table 3-4 Fama and MacBeth Regression Results with the Illiquidity Measure

This table reports the time-series averages of the monthly estimated coefficients in cross-sectional regressions for NYSE and Amex firms, as well as for Nasdaq firms, using the weighted least-square methodology suggested by Litzenberger and Ramaswamy (1979). The dependent variable is the percentage monthly return. *ECIN* is the information risk measure estimated yearly following the model of equation (1.1). *SIZE* is the logarithm of the year-end market value of equity (in million dollars). *BM* is the logarithm of the book value divided by the market value. *BETA* is the post-ranking portfolio betas of Fama-French (1992) estimated using 40 portfolios. *ILLIQ* is the Amihud (2002) illiquidity measure. *ADUM* is equal to one if the firm has one or more analysts, and zero otherwise. *CDUM* is equal to one if the firm has a credit rating, and zero otherwise. *SDUM* is equal to one if the fraction of shares held by large institutions exceeds the median level for the entire sample, and zero otherwise. Panel A reports the Fama-MacBeth regressions for NYSE/Amex common stocks in years 1984-2007, where the regressions involving *ADUM* and *CDUM* start from 1985. Panel B reports the Fama-MacBeth regressions for Nasdaq common stocks in years 1988-2006. The *t*-statistics are in the parentheses.

Panel A NYSE/Amex

	ECIN	Beta	SIZE	BM	ILLIQ	ECIN×ADUM	ECIN×CDUM	ECIN×SDUM
(1)	0.422 (3.10)	-0.418 (-1.69)	0.150 (1.85)	0.117 (3.32)	0.014 (0.28)			
(2)	1.000 (1.99)	-0.441 (-1.78)	0.142 (1.76)	0.106 (3.03)	0.012 (0.26)	-0.028 (-0.06)		
(3)	0.859 (4.46)	-0.42 (-1.7)	0.136 (1.72)	0.107 (2.98)	0.010 (0.21)		-0.431 (-2.28)	
(4)	0.837 (4.41)	-0.453 (-1.85)	0.152 (1.88)	0.109 (3.15)	0.012 (0.24)			-0.404 (-2.28)

Panel B Nasdaq

	ECIN	Beta	SIZE	BM	ILLIQ	ECIN×ADUM	ECIN×CDUM	ECIN×SDUM
(1)	0.893 (2.52)	-0.413 (-1.28)	0.383 (3.09)	0.141 (2.32)	-0.011 (-0.78)			
(2)	1.045 (2.50)	-0.424 (-1.33)	0.376 (3.01)	0.135 (2.02)	-0.01 (-0.72)	-0.412 (-0.75)		
(3)	0.879 (2.52)	-0.414 (-1.29)	0.386 (3.17)	0.145 (2.3)	-0.01 (-0.74)		-0.111 (-0.39)	
(4)	1.017 (2.63)	-0.426 (-1.33)	0.378 (3.05)	0.122 (1.94)	-0.012 (-0.85)			-0.246 (-0.9)

Table 3-5 Fama and MacBeth Regression Results with *ECIN*-Squared

This table reports the time-series averages of the monthly estimated coefficients in cross-sectional regressions for NYSE and Amex firms, as well as for Nasdaq firms, using the weighted least-square methodology suggested by Litzenberger and Ramaswamy (1979). The dependent variable is the percentage monthly return. *ECIN* is the information risk measure estimated yearly following the model of (1.1). *SIZE* is the logarithm of the year-end market value of equity (in million dollars). *BM* is the logarithm of the book value divided by the market value. *BETA* is the post-ranking portfolio betas of Fama-French (1992) estimated using 40 portfolios. *ADUM* is equal to one if the firm has one or more analysts, and zero otherwise. *CDUM* is equal to one if the firm has a credit rating, and zero otherwise. *SDUM* is equal to one if the fraction of shares held by large institutions exceeds the median level for the entire sample, and zero otherwise. Panel A reports the Fama-MacBeth regressions for NYSE/Amex common stocks in years 1984-2007, where the regressions involving *ADUM* and *CDUM* start from 1985. Panel B reports the Fama-MacBeth regressions for Nasdaq common stocks in years 1988-2007. The *t*-statistics are in the parentheses.

Panel A NYSE/Amex

	ECIN	Beta	SIZE	BM	ECIN ²	ECIN×ADUM	ECIN×CDUM	ECIN×SDUM
(1)	0.810 (2.94)	-0.426 (-1.71)	0.149 (1.83)	0.116 (3.47)	0.068 (0.60)			
(2)	1.479 (2.66)	-0.445 (-1.79)	0.140 (1.73)	0.103 (3.05)	0.078 (0.66)	-0.090 (-0.19)		
(3)	1.08 (3.67)	-0.422 (-1.70)	0.133 (1.67)	0.107 (3.06)	0.074 (0.60)		-0.479 (-2.40)	
(4)	1.134 (3.94)	-0.457 (-1.86)	0.151 (1.86)	0.109 (3.30)	0.025 (0.20)			-0.391 (-2.18)

Panel B Nasdaq

	ECIN	Beta	SIZE	BM	ECIN ²	ECIN×ADUM	ECIN×CDUM	ECIN×SDUM
(1)	1.448 (2.03)	-0.386 (-1.21)	0.370 (3.02)	0.162 (2.86)	-0.174 (-0.63)			
(2)	2.018 (2.72)	-0.407 (-1.28)	0.363 (2.94)	0.147 (2.41)	-0.072 (-0.24)	-0.495 (-0.91)		
(3)	1.456 (2.03)	-0.386 (-1.21)	0.371 (3.11)	0.165 (2.88)	-0.166 (-0.58)		-0.105 (-0.37)	
(4)	1.542 (2.18)	-0.397 (-1.24)	0.365 (2.97)	0.141 (2.46)	-0.254 (-0.89)			-0.272 (-0.98)

Table 3-6 Changes in *ECIN* and Information Dissemination

This table reports the change in *ECIN* and the proxies for information dissemination around Regulation FD. *ECIN* is the information risk measure estimated yearly following the model of equation (1.1). *ADUM* is equal to one if the firm has one or more analysts, and zero otherwise. *CDUM* is equal to one if the firm has a credit rating, and zero otherwise. *SDUM* is equal to one if the fraction of shares held by large institutions exceeds the median level for the entire sample, and zero otherwise. Change in *ECIN* is estimated with the vector error-correction model (VECM) run in the period October 23, 1999 to January 23, 2001, with a dummy variable that takes value of one in the period after the effective date of Regulation FD (October 23, 2000) and zero otherwise. *ADUM*, *CDUM* and *SDUM* are the changes from the quarter before Regulation FD to the quarter after Regulation FD. The average changes in *ECIN*, *ADUM*, *CDUM* and *SDUM* are displayed with the corresponding *t*-statistics in parenthesis.

Exchange	Number of firms	Average Changes in			
		<i>ECIN</i>	<i>ADUM</i>	<i>CDUM</i>	<i>SDUM</i>
NYSE/Amex	1519	-0.0282	0.0059	0.0007	0.0020
		(-5.19)	(1.41)	(0.28)	(0.34)
Nasdaq	2388	0.0192	0.0084	0.0021	-0.0013
		(3.79)	(2.24)	(1.39)	(-0.26)

Table 3-7 Firm Level Cross-Sectional Cost of Capital Changes

This table reports the estimated coefficients of firm-level cross-section regressions. The dependent variable is the annualized percentage change in firms' average cost of capital resulting from changes in information asymmetry and information dissemination. The change in firms' average cost of capital is calculated using the Fama-MacBeth average regression coefficients from regression (1) in Table 3-3 multiplied by the changes in *ECIN*, or from regression (2) in Table 3-3 multiplied by the changes in *ECIN* and the analyst dummy (*ADUM*), or from regression (3) in Table 3-3 multiplied by the changes in *ECIN* and the credit rating dummy (*CDUM*), or from Regression (4) in Table 3-3 multiplied by the changes in *ECIN* and institutional share dummy (*SDUM*). The explanatory variable ΔVOL is the change in firms' average monthly trading volume, from the period three months before Regulation FD to the period three months after Regulation FD, expressed in units of ten million shares. The explanatory variable $\Delta LARGE$ is the change in firms' average number of trades with ten thousand shares or more each month, from the period three months before Regulation FD to the period three months after Regulation FD, expressed in units of one hundred trades. *SIZE* is the logarithm of the market value of equity at the end of the year 1999. *BM* is the logarithm of the ratio of the book equity value to the market value of equity at the end of the year 1999. *ECIN* is estimated with the vector error-correction model (VECM) over the year 1999. *HIGHTECH* equals to one if the firm is involved in a high technology industry. All the explanatory variables except dummies are demeaned. Panel A reports the results for NYSE/Amex firms and Panel B reports the results for Nasdaq firms. The *t*-statistics are in the parentheses.

Panel A NYSE/Amex

	Constant	SIZE	BM	ECIN	Δ VOL	Δ LARGE	HITECH	Adj. R ²
ECIN	-0.150 (-5.19)							
	-0.141 (-6.22)	0.008 (0.40)	0.009 (0.30)	0.137 (0.59)				-0.002
	-0.133 (-5.41)	-0.017 (-0.83)	0.007 (0.21)	0.055 (0.23)	-0.005 (-0.08)	1.909 (0.69)	0.035 (0.41)	0.001
	-0.334 (-5.10)							
ADUM	-0.312 (-6.20)	0.019 (0.44)	0.020 (0.29)	0.301 (0.58)				-0.002
	-0.292 (-5.38)	-0.039 (-0.84)	0.014 (0.19)	0.118 (0.22)	-0.012 (-0.08)	4.193 (0.68)	0.08 (0.43)	0.001
	-0.224 (-4.54)							
	-0.198 (-6.04)	0.032 (1.13)	0.020 (0.44)	0.197 (0.59)				-0.001
CDUM	-0.171 (-5.03)	-0.018 (-0.60)	0.009 (0.20)	0.058 (0.17)	-0.013 (-0.14)	2.707 (0.70)	0.007 (0.06)	-0.001
	-0.205 (-4.38)							
	-0.186 (-6.23)	0.029 (1.12)	0.019 (0.46)	0.184 (0.60)				-0.001
	-0.159 (-5.13)	-0.023 (-0.85)	0.000 (0.01)	0.074 (0.24)	-0.075 (-0.91)	4.450 (1.27)	0.057 (0.53)	-0.001
SDUM								

Panel B Nasdaq

	Constant	SIZE	BM	ECIN	Δ VOL	Δ LARGE	HITECH	Adj. R ²
ECIN	0.204 (3.79)							
	0.179 (3.12)	-0.309 (-6.10)	-0.275 (-3.99)	-2.275 (-5.14)				0.024
	0.159 (2.25)	-0.412 (-8.18)	-0.226 (-3.44)	-2.466 (-6.13)	0.136 (2.24)	-9.540 (-2.00)	0.226 (1.77)	0.045
ADUM	0.180 (3.50)							
	0.161 (2.94)	-0.253 (-5.22)	-0.239 (-3.62)	-1.696 (-4.00)				0.016
	0.158 (2.63)	-0.349 (-8.17)	-0.189 (-3.38)	-1.871 (-5.48)	0.104 (2.02)	-7.09 (-1.75)	0.183 (1.69)	0.041
CDUM	0.201 (3.82)							
	0.177 (3.17)	-0.302 (-6.11)	-0.269 (-4.01)	-2.198 (-5.10)				0.024
	0.156 (2.28)	-0.401 (-8.23)	-0.221 (-3.46)	-2.38 (-6.11)	0.134 (2.28)	-9.516 (-2.06)	0.225 (1.82)	0.045
SDUM	0.242 (4.18)							
	0.215 (3.51)	-0.317 (-5.83)	-0.303 (-4.11)	-2.309 (-4.88)				0.023
	0.219 (2.99)	-0.427 (-8.17)	-0.258 (-3.78)	-2.482 (-5.95)	0.125 (1.99)	-8.536 (-1.73)	0.191 (1.44)	0.044

Table 3-8 Cost of Capital Changes and Firm Characteristics

The table reports the estimated coefficients of firm-level cross-section regressions. The dependent variable is the annualized percentage change in firms' average cost of capital resulting from changes in information asymmetry and information dissemination. The change in firms' average cost of capital is calculated using the Fama-MacBeth average Regression coefficients from regression (1) in Table 3-3 multiplied by the changes in *ECIN*. *SIZE* is the logarithm of the market value of equity at the end of year 1999. *BM* is the logarithm of the ratio of book equity value to the market value of equity at the end of the year 1999. *ECIN* is estimated with the vector error-correction model (VECM) over the year 1999. *NASDAQ*, a dummy variable equals to one if the firm is in Nasdaq and zero if the firm is in NYSE/Amex. All the explanatory variables except dummies are demeaned. The *t*-statistics are in the parentheses.

	Constant	NASDAQ	SIZE	BM	ECIN	Adj. R ²
ECIN	-0.141 (-2.46)	0.320 (4.33)				0.005
	0.024 (0.38)	0.047 (0.54)	-0.191 (-6.33)	-0.185 (-4.18)	-1.713 (-5.70)	0.020
ADUM	-0.312 (-5.05)	0.474 (5.95)				0.010
	-0.177 (-2.59)	0.250 (2.67)	-0.151 (-4.64)	-0.156 (-3.27)	-1.240 (-3.82)	0.017
CDUM	-0.198 (-3.42)	0.375 (5.03)				0.007
	-0.048 (-0.75)	0.126 (1.44)	-0.174 (-5.71)	-0.174 (-3.91)	-1.636 (-5.39)	0.020
SDUM	-0.186 (-2.98)	0.401 (5.00)				0.007
	-0.023 (-0.33)	0.131 (1.39)	-0.186 (-5.67)	-0.199 (-4.14)	-1.738 (-5.32)	0.021

REFERENCES

- Aboody, D., J. Hughes, and J. Liu, 2005, Earning quality, insider trading and cost of capital, *Journal of Accounting Research* 43, 651-673.
- Admati, A., 1985, A noisy rational expectations equilibrium for multi-asset securities markets, *Econometrica* 53, 629-658.
- Admati, A., and P. Pfleiderer, 1988, A theory of intraday patterns: volume and price variability, *Review of Financial Studies* 1, 3-40.
- Agrawal, Anup, Sahiba Chadha, and Mark A. Chen, 2006, Who is afraid of Reg FD? The behavior and performance of sell-side analysts following the sec's fair disclosure rules, *Journal of Business* 79, 2811-2834.
- Aktas, Nihat, Eric de Bodt, Fany Declerck, and Herve Van Oppens, 2007, The PIN anomaly around M&A announcements, *Journal of Financial Markets* 10, 169-191.
- Almazan, Andres, Javier Suarez, and Sheridan Titman, 2003, Stakeholder, transparency and capital structure, Working paper, University of Texas at Austin and European Corporate Governance Institute.
- Amihud, Yakov, 2002, Illiquidity and stock returns: cross-section and time-series effects, *Journal of Financial Markets* 5, 31-56.
- Amihud, Yakov, and Haim Mendelson, 1986, Asset pricing and the bid-ask spread, *Journal of Financial Economics* 17, 223-249.
- Ang, A., R. J. Hordrick, Y. Xing, and X. Zhang, 2006, The cross-section of volatility and expected return, *The Journal of Finance* 73, 289-231.
- Ascioglu, Asli, Shantaram P. Hegde, and John B. McDermott, 2008, Information asymmetry and investment-cash flow sensitivity, *Journal of Banking & Finance* 32, 1036-1048.
- Aslan, Hadiye, David Easley, Soeren Hvidkjaer, and Maureen O'Hara, 2007, Firm characteristics and informed trading: Implications for asset pricing, Working paper, University of Houston, Cornell University, and Copenhagen Business School.
- Bacidore, Jeffrey M., 1997, The impact of decimalization on market quality: an empirical investigation of the Toronto Stock Exchange, *Journal of Financial Intermediation* 6, 92-120.
- Bacidore Jeffrey M., 2001, Decimalization, adverse selection, and market maker rents, *Journal of Banking and Finance* 25, 829-855.
- Bailey, Warren, Haitao Li, Connie X. Mao, and Rui Zhong, 2003, Regulation fair disclosure and earnings information: market, analyst, and corporate responses, *The Journal of Finance* 58, 2487-2514.
- Ball, R., and R. Brown, 1968, An empirical evaluation of accounting numbers, *Journal of Accounting Research* 6, 159-178.
- Barberis, N., A. Shleifer, and R. Vishny, 1998, A model of investor sentiment, *Journal of Financial Economics* 49, 307-343.
- Barclay, Michael J., and Jerold B. Warner, 1993, Stealth trading and volatility: which trades move prices?, *Journal of Financial Economics* 34, 281-305.
- Barry, C., and S. J. Brown, 1985, Differential information and security market equilibrium, *Journal of Financial and Quantitative Analysis* 20, 407-422.
- Battalio, Robert H., and Richard R. Mendenhall, 2005, Earnings expectations, investor trade size, and anomalous returns around earnings announcement, *Journal of Financial Economics* 77, 289-319.
- Bernard, V. L., and J. K. 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.

- Black, Fischer, Michael C. Jensen, and Myron S. Scholes, 1972, The capital asset pricing model: some empirical tests, in Michael Jensen, ed.: *Studies in the theory of capital market* (Praeger, New York).
- Booth, G. Geoffrey, Ji-Chai Lin, Teppo Martikainen, and Yiuman Tse, 2002, Trading and pricing in upstairs and downstairs stock markets, *The Review of Financial Studies* 15, 1111-1135.
- Botosan, C., 1997, Disclosure level and the cost of equity capital, *The Accounting Review* 72, 323-349.
- Botosan, C. A., and M. A. Plumlee, 2002, A re-examination of disclosure level and the expected cost of equity capital, *Journal of Accounting Research* 40, 21-40.
- Botosan, C., M. Plumlee, and Y. Xie, 2004, The role of private information precision in determining cost of equity capital, *Review of Accounting Studies* 9, 233-259.
- Brav, Alon, Christopher Geczy, and Paul A. Gompers, 2000, Is the abnormal return following equity issuances anomalous?, *Journal of Financial Economics* 56, 209-249.
- Brav, Alon, and Paul A. Gompers, 1997, Myth or reality? The long-run underperformance of initial public offerings: evidence from venture and nonventure capital-backed companies, *The Journal of Finance* 52, 1791-1821.
- Brav, Alon, and J. B. Heaton, 2002, Competing theories of financial anomalies, *The Review of Financial Studies* 15, 575-606.
- Brennan, M., and A. Subrahmanyam, 1995, Investment analysis and price formation in securities markets, *Journal of Financial Economics* 38, 361-381.
- Brennan, M.J., A. Wang, and Y. Xiao, 2004, Estimation and test of a simple model of intertemporal capital asset pricing, *The Journal of Finance* 59, 1743-1776.
- Brennan, Michael J., and Avanidhar Subrahmanyam, 1996, Market microstructure and asset pricing: on the compensation for illiquidity in stock returns, *Journal of Financial Economics* 41, 441-464.
- Brown, Stephen, Stephen A. Hillegeist, and Kin Lo, 2004, Conference calls and information asymmetry, *Journal of Accounting and Economics* 37, 343-366.
- Bushee, Brain J., Dawn A. Matsumoto, and Gregory S. Miller, 2004, Managerial and investor responses to disclosure regulation: the case of Reg FD and conference call, *The Accounting Review* 79, 617-643.
- Campbell, J.Y., and T. Vuolteenaho, 2004, Bad beta, good beta, *American Economic Review* 94, 1249-1275.
- Carhart, Mark M., 1997, On persistence in mutual fund performance, *The Journal of Finance* 52, 57-82.
- Carlson, Murray, Adlai Fisher, and Ron Giammarino, 2006, Corporate investment and asset price dynamics: implications for SEO event studies and long-run performance, *The Journal of Finance* 61, 1009-1034.
- Caskey, Judson A., 2008, Information in equity markets with ambiguity-averse investors, *Review of Financial Studies* Forthcoming.
- Chakravarty, Sugato, 2001, Stealth-trading: which trader's trades move stock prices?, *Journal of Financial Economics* 61, 289-307.
- Chakravarty, Sugato, Bonnie F. Van Ness, and Robert A. Van Ness, 2005, The effect of decimalization on trade size and adverse selection costs, *Journal of Business Finance & Accounting* 32, 1063 – 1081.
- Chan, Kalok, and Wai-Ming Fong, 2000, Trade size, order imbalance, and the volatility-volume relation, *Journal of Financial Economics* 57, 247-273.
- Chan, Louis K. C., Narasimhan Jegadeesh, and Josef Lakonishok, 1996, Momentum strategies, *The Journal of Finance* 51, 1681-1713.
- Chan, Wesley S., 2003, Stock price reaction to news and no-news: drift and reversal after headlines, *Journal of Financial Economics* 70, 223-260.

- Chen, Qi, Itay Goldstein, and Wei Jiang, 2007, Price informativeness and investment sensitivity to stock price, *Review of Financial Studies* 20, 619-650.
- Chordia, Tarun, and Lakshmanan Shivakumar, 2006, Earnings and price momentum, *Journal of Financial Economics* 80, 627-656.
- Chordia, Tarun, Avanidhar Subrahmanyam, and V. R. Anshuman, 2001, Trading activity and expected stock returns, *Journal of Financial Economics* 59, 3-32.
- Chou, Robin K., and Huimin Chung, 2004, Decimalization, trading costs, and information transmissions between ETFs and index futures, Working paper, National Central University and National Chiao-Tung University.
- Chung, Kee H., and Xin Zhao, 2005, Decimal pricing and information-based trading: tick size and informational efficiency of asset price, Working paper, SUNY at Buffalo and Pennsylvania State University (Erie).
- Clarkson, P., J. Guedes, and R. Thompson, 1996, On the diversification, observability, and measurement of estimation risk, *Journal of Financial and Quantitative Analysis* 31, 69-84.
- Cohen, Randolph B., Paul A. Gompers, and Tuomo Vuolteenaho, 2002, Who underreacts to cash-flow news? Evidence from trading between individuals and institutions, *Journal of Financial Economics* 66, 409-462.
- Coles, Jeffrey L., Uri Loewenstein, and Jose Suay, 1995, On equilibrium pricing under parameter uncertainty, *The Journal of Financial and Quantitative Analysis* 30, 347-364.
- Collver, Charles D., 2007, Is there less informed trading after Regulation Fair Disclosure, *Journal of Corporate Finance* 13, 270-281.
- Comment, R., and G. A. Jarrell, 1991, The relative signaling power of Dutch-auction and fixed-price self-tender offers and open-market share repurchases, *The Journal of Finance* 46, 1243-1271.
- Core, John E., Wayne R. Guay, and Rodrigo Verdi, 2008, Is accruals quality a priced risk factor?, *Journal of Accounting and Economics* 46, 2-22.
- D'Mello, Ranjan, Stephen P. Ferris, and Chuan Yang Hwang, 2003, The tax-loss selling hypothesis, market liquidity, and price pressure around the turn-of-the-year, *Journal of Financial Markets* 6, 73-98.
- Daniel, Kent, David Hirshleifer, and Avanidhar Subrahmanyam, 1998, Investor psychology and security market under- and overreactions, *The Journal of Finance* 53, 1839-1885.
- Dann, L. Y., 1981, Common stock repurchases: an analysis of returns to bondholders and stockholders, *Journal of Financial Economics* 9, 113-138.
- Diamond, Douglas W., and Robert E. Verrecchia, 1987, Constraints on short-selling and asset price adjustment to private information, *Journal of Financial Economics* 18, 277-311.
- Diamond, Douglas W., and Robert E. Verrecchia, 1991, Disclosure, liquidity, and the cost of capital, *The Journal of Finance* 46, 1325-1359.
- Duarte, Jefferson, Xi Han, Jarrad Harford, and Lance Young, 2008, Information asymmetry, information dissemination and the effect of Regulation FD on the cost of capital, *Journal of Financial Economics* 87, 24-44.
- Duarte, Jefferson, and Lance Young, 2009, Why is PIN priced?, *Journal of Financial Economics* 91, 119-138.
- Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2002, Is information risk a determinant of asset returns?, *The Journal of Finance* 57, 2185-2221.
- Easley, David, Soeren Hvidkjaer, and Maureen O'Hara, 2005, Factoring information into returns, Working paper, Cornell University and Copenhagen Business School.
- Easley, David, Nicholas M. Kiefer, and Maureen O'Hara, 1997, The information content of the trading process, *Journal of Empirical Finance*, 159-186.
- Easley, David, Nicholas M. Kiefer, Maureen O'Hara, and Joseph B. Paperman, 1996, Liquidity, information, and infrequently traded stocks, *The Journal of Finance* 51, 1405-1436.
- Easley, David, and Maureen O'Hara, 1987, Price, trade size, and information in securities markets, *Journal of Financial Economics* 19, 69-90.

- Easley, David, and Maureen O'Hara, 2004, Information and the cost of capital, *The Journal of Finance* 59, 1553-1583.
- Eckbo, B. Espen, Ronald W. Masulis, and Oyvind Norli, 2000, Seasoned public offerings: resolution of the 'new issues puzzle', *Journal of Financial Economics* 56, 251-291.
- Eleswarapu, Venkat R., Rex Thompson, and Kumar Venkataraman, 2004, The impact of Regulation Fair Disclosure: trading costs and information asymmetry, *The Journal of Financial and Quantitative Analysis* 39, 209-225.
- Engle, Robert F., and Clive W. J. Granger, 1987, Cointegration and error correction: representation, estimation and testing, *Econometrica* 55, 251-276.
- Eun, Cheol S., and Sanjiv Sabherwal, 2003, Cross-border listings and price discovery: evidence from U.S.-listed Canadian stocks, *The Journal of Finance* 58, 549-575.
- Fama, Eugene F., 1976. *Foundations of finance* (Basic Books, New York).
- Fama, Eugene F., and James D. MacBeth, 1973, Risk, return and equilibrium: empirical tests, *Journal of Political Economy* 81, 607-636.
- Fama, Eugene F., 1998, Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* 49, 283-306.
- Fama, Eugene F., and Kenneth R. French, 1992, The cross-section of expected stock returns, *The Journal of Finance* 47, 427-465.
- Fama, Eugene F., and Kenneth R. French, 1993, Common risk factors in the returns on stocks and bonds, *Journal of Financial Economics* 33, 3-56.
- Fama, Eugene F., and Kenneth R. French, 2008, Average returns, B/M and share issues, *The Journal of Finance* 63, 2971-2995.
- Fama, Eugene F., and James D. MacBeth, 1973, Risk, return, and equilibrium: empirical tests, *Journal of Political Economy* 81, 607-636.
- Finucane, Tomas J., 2000, A direct test of methods for inferring trade direction from intra-day data, *Journal of Financial and Quantitative Analysis* 35, 553-576.
- Ferreira, Miguel A., and Paul A. Laux, 2007, Corporate governance, idiosyncratic risk, and information flow, *The Journal of Finance* 62, 951-989.
- Francis, Jennifer, Ryan Lafond, Per Olsson, and Katherine Schipper, 2004, Costs of capital and earnings attributes, *The Accounting Review* 79, 967-1010.
- Francis, Jennifer, Ryan Lafond, Per Olsson, and Katherine Schipper, 2007, Information uncertainty and post-earnings-announcement-drift, *Journal of Business Finance & Accounting* 34, 403-433.
- Francis, Jennifer, Dhananjay Nanda, and Per Olsson, 2008, Voluntary disclosure, earnings quality, and cost of capital, *Journal of Accounting Research* 46, 53-99.
- Garleanu, Nicolae, and Lasse Heje Pedersen, 2004, Adverse selection and the required return, *The Review of Financial Studies* 17, 643-665.
- George, Thomas J., Gautam Kaul, and Mahendrarajah Nimalendran, 1991, Estimation of the bid-ask spread and its components: a new approach, *Review of Financial Studies* 4, 623-656.
- George, Thomas J., and Chuan Yang Hwang, 2004, The 52-week high and momentum investing, *The Journal of Finance* 59, 2145-2176.
- George, Thomas J., and Chuan Yang Hwang, 2007, Long-term return reversals: overreaction or taxes? , *The Journal of Finance* 62, 2865 - 2896.
- Gibbons, Michael R., Stephen A. Ross, and Jay Shanken, 1989, A test of the efficiency of a given portfolio, *Econometrica* 57, 1121-1152.
- Gibson, Scott, Rajdeep Singh, and Vijay Yerramilli, 2003, The effect of decimalization on the components of the bid-ask spread, *Journal of Financial Intermediation* 12, 121-148.
- Gintschel, Andreas, and Stanimir Markov, 2004, The effectiveness of Regulation FD, *Journal of Accounting and Economics* 37, 293-314.

- Glosten, Lawrence R., and Paul R. Milgrom, 1985, Bid, ask, and transaction prices in a specialist market with heterogeneously informed traders, *Journal of Financial Economics* 14, 71-100.
- Glosten, Lawrence R., and Lawrence E. Harris, 1988, Estimating the components of the bid/ask spread, *Journal of Financial Economics* 21, 123-142.
- Gomes, Armando, Gary Gorton, and Leonardo Madureira, 2007, SEC Regulation Fair Disclosure, information, and the cost of capital, *Journal of Corporate Finance* 13, 300-334.
- Grinblatt, Mark, and Tobias J. Moskowitz, 2004, Predicting stock price movements from past returns: the role of consistency and tax-loss selling, *Journal of Financial Economics* 71, 541-579.
- Grossman, Sanford J., 1976, On the efficiency of competitive stock markets where trades have diverse information, *The Journal of Finance* 31, 573-585.
- Grossman, Sanford J., and Joseph E. Stiglitz, 1980, On the impossibility of informationally efficient markets, *The American Economic Review* 70, 393-408.
- Grundy, B., and McNichols M., 1989, Trade and revelation of information through prices and direct disclosure, *Review of Financial Studies* 2, 495-526.
- Handa, Punnett, and Scott C. Linn, 1993, Arbitrage pricing with estimation risk, *Journal of Financial and Quantitative Analysis* 28, 81-100.
- Harris, Frederick H. deB., Thomas H. McInish, Gary L. Shoesmith, and Robert A. Wood, 1995, Cointegration, error correction, and price discovery on informationally linked security markets, *The Journal of Financial and Quantitative Analysis* 30, 563-578.
- Hasbrouck, Joel, 1991a, Measuring the information content of stock trades, *Journal of Finance* 46, 179-207.
- Hasbrouck, Joel, 1991b, The summary informativeness of stock trades: an econometric analysis, *Review of Financial Studies* 4, 571-595.
- Hasbrouck, Joel, 1995, One security, many markets: determining the contribution to price discovery, *The Journal of Finance* 50, 1175-1199.
- Hausch, D. B., and J. K. Seward, 1993, Signaling with dividends and share repurchases: a choice between deterministic and stochastic cash disbursements, *Review of Financial Studies* 6, 121-154.
- Hausman, J. A., A. W. Lo, and A. C. MacKinlay, 1992, An ordered probit analysis of transaction stock prices, *Journal of Financial Economics* 31, 319-379.
- Heflin, Frank, K. R. Subramanyam, and Yuan Zhang, 2003, Regulation FD and the financial information environment: early evidence, *The Accounting Review* 78, 1-37.
- Hendershott, Terrence, and Charles M. Jones, 2005, Island goes dark: transparency, fragmentation, and regulation, *The Review of Financial Studies* 18, 743-793.
- Hirshleifer, David A., Kewei Hou, and Siew Hong Teoh, 2006, The accrual anomaly: risk or mispricing?, Working paper, University of California, Irvine and Ohio State University.
- Holthausen, Robert W., Richard W. Leftwich, and David Mayers, 1987, The effects of large block transaction on security prices: a cross-sectional analysis, *Journal of Financial Economics* 19, 237-267.
- Holthausen, Robert W., Richard W. Leftwich, and David Mayers, 1990, Large-block transactions, the speed of response, and temporary and permanent stock-price effects, *Journal of Financial Economics* 26, 71-95.
- Holthausen, Robert W., and Robert E. Verrecchia, 1990, The effect of informedness and consensus on price and volume behavior, *Accounting Review* 65, 191-208.
- Hong, Harrison, Terence Lim, and Jeremy C. Stein, 2000, Bad news travels slowly: size, analyst coverage, and the profitability of momentum strategies, *The Journal of Finance* 55, 265-295.
- Hong, Harrison, and Jeremy C. Stein, 1999, A unified theory of underreaction, momentum trading, and overreaction in asset markets, *The Journal of Finance* 54, 2143-2184.
- Huang, R. D., and H. R. Stoll, 1997, The components of the bid-ask spread: a general approach, *Review of Financial Studies* 10, 995-1034.

- Huang, Roger D., 2002, The quality of ECN and NASDAQ market maker quotes, *The Journal of Finance* 57, 1285-1319.
- Hughes, John S, Jing Liu, and Jun Liu, 2007, Information asymmetry, diversification, and cost of capital, *The Accounting Review* 82, 705-730.
- Hvidkjaer, Soeren, 2006, A trade-based analysis of momentum, *The Review of Financial Studies* 19, 457-491.
- Irani, Afshad J., and Irene Karamanou, 2003, Regulation Fair Disclosure, analyst following, and analyst forecast dispersion, *Accounting Horizons* 17, 15-29.
- Jagannathan, Ravi, and Zhenyu Wang, 1996, The conditional CAPM and the cross-section of expected returns, *The Journal of Finance* 51, 3-53.
- Jayaraman, Sudarshan, 2008, Earnings volatility, cash flow volatility and informed trading, *Journal of Accounting Research* 46, 809-851.
- Jegadeesh, Narasimhan, and Sheridan Titman, 1993, Returns to buying winners and selling losers: implications for stock market efficiency, *The Journal of Finance* 48, 65-91.
- Johansen, Soren, 1988, Statistical analysis of cointegration vectors, *Journal of Economic Dynamics and Control* 12, 231-254.
- Jorion, Philippe, Zhu Liu, and Charles Shi, 2005, Informational effects of Regulation FD: evidence from rating agencies, *Journal of Financial Economics* 76, 309-330.
- Ke, Bin, Kathy Petroni, and Yong Yu, 2008, The effect of Regulation FD on transient institutional investors' trading behavior, *Journal of Accounting Research* 46, 853-883.
- Kim, O., and R. Verrecchia, 1991, Market reactions to anticipated announcements, *Journal of Financial Economics* 30, 273-310.
- Kempf, A., and O. Korn, 1999, Market depth and order size, *Journal of Financial Markets* 2, 29-48.
- Khan, M., 2008, Are accruals mispriced? Evidence from tests of an intertemporal capital asset pricing model., *Journal of Accounting and Economics* 45, 55-77.
- Kyle, Albert S., 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315-1336.
- LaFond, Ryan, and Ross L. Watts, 2008, The information role of conservatism, *Accounting Review* 83, 447-478.
- Lambert, R., C. Leuz, and R. Verrecchia, 2007, Accounting information, disclosure, and cost of capital, *Journal of Accounting Research* 45, 385-420.
- Lambert, Richard, Christian Leuz, and Robert E. Verrecchia, 2006, Accounting information, disclosure, and the cost of capital, Working paper, University of Pennsylvania and University of Chicago.
- Lang, Mard H., and Russell J. Lundholm, 2000, Voluntary disclosure and equity offerings: reducing information asymmetry or hyping the stock?, *Contemporary Accounting Research* 17, 623-662.
- Lee, C. M., 1992, Earnings news and small traders, *Journal of Accounting and Economics* 15, 265-302.
- Lee, C. M., B. Mucklow, and M. J. Ready, 1993, Spreads, depths, and the impact of earnings information: an intraday analysis, *Review of Financial Studies* 6, 345-374.
- Lee, C. M., and B. Radhakrishna, 2000, Inferring investor behavior: evidence from TORQ data, *Journal of Financial Markets* 3, 83-111.
- Lee, C. M., and M. J. Ready, 1991, Inferring trade direction from intraday data, *The Journal of Finance* 55, 2017-2069.
- Levi, Maurice D., 1973, Errors in the variables bias in the presence of correctly measured variables, *Econometrica* 41, 985-986.
- Lin, J., G. Sanger, and G. G. Booth, 1995, Trade size and components of the bid-ask spread, *Review of Financial Studies* 8, 1153-1183.

- Litzenberger, Robert, and Krishna Ramaswamy, 1979, The effect of personal taxes and dividends on capital asset prices: theory and empirical evidence, *Journal of Financial Economics* 7, 163-196.
- Liu, Michelle, and Peter Wysocki, 2006, Operating risk, information risk, and cost of capital, Working paper, Pennsylvania State University and MIT Sloan School of Management.
- Livnat, Joshua, and Richard R. Mendenhall, 2006, Comparing the post-earning announcement drift for surprises calculated from analyst and time series forecasts, *Journal of Accounting Research* 44, 177-205.
- Llorente, Guillermo, Roni Michaely, Gideon Saar, and Jiang Wang, 2002, Dynamic volume-return relation of individual stocks, *The Review of Financial Studies* 15, 1005-1047.
- Lo, Andrew W., and A. Craig MacKinlay, 1990, An econometric analysis of nonsynchronous trading, *Journal of Econometrics* 45, 181-211.
- Loughran, Tim, and Jay R. Ritter, 1995, The new issues puzzle, *The Journal of Finance* 50, 23-51.
- Madhavan, A., M. Richardson, and M. Roomans, 1997, Why do security prices change? A transaction-level analysis of nyse stocks, *Review of Financial Studies* 10, 1035-1064.
- Madhavan, A., and S. Smidt, 1991, A Bayesian model of intraday specialist pricing, *Journal of Financial Economics* 30, 99-134.
- Masulis, R. W., 1980, Stock repurchase by tender offer: an analysis of the causes of common stock price changes, *The Journal of Finance* 35, 309-319.
- Mohanram, Partha, and Shiva Rajgopal, 2009, Is PIN priced risk?, *Journal of Accounting and Economics* 47, 226-243.
- Mohanram, Partha, and Shyam Sunder, 2006, How has Regulation FD affected the operations of financial analysts?, *Contemporary Accounting Research* 23, 491-525.
- Ng, J., T. Rusticus, and R. Verdi, 2007, Implications of trading costs for post-earnings-announcement drift, *Journal of Accounting Research* 46, 661-696.
- O'Hara, Maureen, 2003, Presidential address: liquidity and price discovery., *The Journal of Finance* 58, 1335-1354.
- Ofer, A. R., and A. V. Thakor, 1987, A theory of stock price responses to alternative corporate cash disbursement methods: stock repurchases and dividends, *The Journal of Finance* 42, 365-394.
- Pástor, Luboš, and Robert F. Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642-685.
- Penman, Stephen H., 1983, The predictive content of earnings forecasts and dividends, *The Journal of Finance* 38, 1181-1199.
- Petkova, R., 2006, Do the Fama-French factors proxy for innovations in predictive variables?, *The Journal of Finance* 61, 581-612.
- Pontiff, Jeffrey, and Artemiza Woodgate, 2008, Share issuance and cross-sectional returns, *The Journal of Finance* 63, 921-945.
- Rangan, Srinivasan, 1998, Earnings management and the performance of seasoned equity offerings, *Journal of Financial Economics* 50, 101-122.
- Roll, R., 1983, Vas ist das? The turn-of-the-year effect and the return premium of small firms, *Journal of Portfolio Management* 9, 18-28.
- Rozeff, Michael S., 1976, Capital market seasonality: the case of stock returns, *Journal of Financial Economics* 3, 379-402.
- Sadka, Ronnie, 2006, Momentum and post-earnings-announcement drift anomalies: the role of liquidity risk, *Journal of Financial Economics* 80, 309-349.
- Schreiber, Paul S., and Robert A. Schwartz, 1986, Price discovery in securities markets, *Journal of Portfolio Management* 12, 43-48.
- Seppi, Duane J., 1990, Equilibrium block trading and asymmetric information, *Journal of Finance* 45, 73-94.
- Shanken, Jay, 1992, On the estimation of beta-pricing models, *The Review of Financial Studies* 5, 1-33.
- Shivakumar, Lakshmanan, 2000, Do firms mislead investors by overstating earnings before seasoned equity offering?, *Journal of Accounting and Economics* 29, 339-371.

- Shleifer, Andrei, and Robert W. Vishny, 1997, The limits of arbitrage, *The Journal of Finance* 52, 35-55.
- Soffer, Leonard C., and Thomas Lys, 1999, Post-earnings announcement drift and the dissemination of predictable information, *Contemporary Accounting Research* 16, 305-331.
- Spiess, D., and J. Affleck-Graves, 1995, Underperformance in long-run stock returns following seasoned equity offerings, *Journal of Financial Economics* 38, 243-267.
- Stock, J. H., and M. W. Watson, 1988, Testing for common trends, *Journal of American Statistical Association* 83, 1097-1107.
- Teoh, Siew Hong, Ivo Welch, and T. J. Wong, 1998, Earnings management and the underperformance of seasoned equity offerings, *Journal of Financial Economics* 50, 63-99.
- Tong, Qing, 2009, Abnormal volume in large trades and the cross-section of expected stock returns, Working paper, Emory University.
- Vega, Clara, 2006, Stock price reaction to public and private information, *Journal of Financial Economics* 82, 103-113.
- Vermaelen, T., 1981, Common stock repurchases and market signalling: an empirical study, *Journal of Financial Economics* 9, 139-183.
- Vuolteenaho, T., 2002, What drives firm-level stock returns, *The Journal of Finance* 57, 233-264.
- Wang, John, 2008. An examination of the impact of voluntary disclosure on post-earnings announcement drift, Working paper, Florida International University.
- Wang, Jiang, 1993, A model of intertemporal asset prices under asymmetric information, *The Review of Economic Studies* 60, 249-282.
- Werner, Ingrid M., and Allan W. Kleidon, 1996, U.K. and U.S. Trading of British cross-listed stocks: an intraday analysis of market integration, *The Review of Financial Studies* 9, 619-664.
- Wooldridge, Jeffrey M., 2001. *Econometric analysis of cross section and panel data* (The MIT Press).