The Interdependence between Institutional Investor Stock Ownership and Information Dissemination by Capital Market Data Aggregators

Julia D'Souza
S.C. Johnson Graduate School of Management
Cornell University
Ithaca, NY 14853
(607) 255-2349
jd48@cornell.edu

K. Ramesh
Eli Broad College of Business
Michigan State University
East Lansing, MI 48824
(517) 432-8350
rameshk@msu.edu

Min Shen
School of Management
George Mason University
Fairfax, VA 22030
(703) 993-1849
mshen3@gmu.edu

Current Version: August 27, 2007

The partial sponsorship of the research by Analysis Group, Inc. is gratefully acknowledged. The authors acknowledge the extensive discussions with and the inputs received from Gary Barwick, Don Barsody, Joe Kannikal, Paul Mindrup, and Debbie Puetz of Standard & Poor's. We also acknowledge the comments and suggestions of Mark Bagnoli, Tom Lys, Karen Nelson, Frank Selto, Jake Thomas, Jim Vincent, Susan Watts, Frank Zhang, and the workshop participants at Penn State, Purdue University, and the University of Colorado at Boulder on an earlier version of the paper. We have benefited immensely from discussions with Scott Eston and Chuck Taylor. We are responsible for all remaining errors. The authors gratefully acknowledge the contributions of I/B/E/S International Inc. for providing analyst forecast data. This data has been provided as part of a broad academic program to encourage earning expectations research.

The Interdependence between Institutional Investor Stock Ownership and Information Dissemination by Capital Market Data Aggregators

Abstract

This paper examines the interdependence between institutional investor stock ownership and the speed with which S&P disseminates corporate accounting information to its commercial customers. From the demand-side perspective, we find that while institutional investors generally influence dissemination speed, quasi-indexers, who rely on corporate accounting information as a low-cost monitoring system, appear to be the key driver of the institutional demand for speedy information dissemination. In addition, dissemination speed increases substantially for stocks listed in major market indices, possibly due to the heightened investor awareness of index stocks. However, data collection lag is longer for stocks with high arbitrage risk or transaction costs, consistent with the documented inability of institutional investors to fully exploit accounting-based mispricing in these circumstances. From the perspective of institutional investor response to dissemination speed, the study finds that both transient investors and quasi-indexers gravitate to stocks with faster information dissemination, consistent with the latter using accounting information as a low cost performance monitoring mechanism, and the former being better enabled to implement their trading strategies in a richer information environment. Overall, the paper provides new insights into the capital market information infrastructure by examining how information intermediaries and sophisticated investors impact each others' resource allocation decisions.

Keywords: *institutional investors; data aggregators; information dissemination; capital markets*

The Interdependence between Institutional Investor Stock Ownership and Information Dissemination by Capital Market Data Aggregators

"The acquisition of information and its dissemination to other economic units are, as we all know, central activities in all areas of finance, and especially so in capital markets." (Merton 1987)

1. Introduction

This paper examines the mutual influence of institutional investor stock ownership and the dissemination speed of key corporate accounting information by Standard & Poor's Compustat (a well-known capital market data aggregator). Academic research and anecdotal evidence show important economic links between a firm's informational characteristics, including its disclosure practices, and key traits such as cost of capital, investor clientele and stock price volatility (see, for instance, Potter 1992; Sias 1996; Botosan 1997; Fox 1997; Serwer 1997, Healy et al. 1999; Bushee and Noe 2000). However, there is no substantial body of research that examines the role played by data aggregators in providing readily accessible corporate financial data in standardized formats to facilitate comparability and valuation, and potentially affect perceived information risk.[1] This paper seeks to address this gap.

The data aggregator we choose to study, Standard & Poor's (S&P, hereafter), is an important information intermediary, which acts as a key supplier in the market for corporate accounting information. The role of data aggregators such as S&P is to design standard data delivery formats, ensure quality control in data collection, and offer database search and screening techniques to mitigate the information overload faced by even sophisticated capital market participants (Ho and Tang 2001). The growing importance of data aggregators is highlighted in the following quote:

[1] See Bushee et al. (2007) for a study of business press as a capital market information intermediary. Within an intra-organizational context, Hansen and Haas (2001) examine information dissemination and use in a management consulting firm.

> "There is clearly a well established and growing multi-billion dollar market for the consumption of high quality financial data as demonstrated by firms such as CompuStat, FactSet, Bloomberg, Thompson Financial, Reuters and others. Each of these data aggregators differentiates itself against its peers by trying to get access to corporate information as fast as it is disseminated to the SEC and from the wire distribution companies, adding value to the data by bundling it with their key search words, filters, analytics, productivity tools and research news so that they gain a higher market share of the data consumers in the industry. Each of these product vendors provides some unique value to its customer base and serves a very valuable function in the efficient operation of today's capital markets." [2] (http://www.wgint.com/ir/niri/be_in_the_know.html)

The sample for our study consists of 312,775 firm-quarters during the period 1991-2004 for which we have S&P's data collection lags, measured as the difference between the periodic SEC filing dates and the corresponding Compustat Research Insight production dates when information from SEC filings were disseminated by S&P to its commercial customers. Descriptive evidence suggests substantial cross-sectional variation in the collection lag during our sample period with a median lag and an inter-quartile range of 15 weekdays each. The large cross-sectional variation provides an impetus for examining the market forces that guide S&P's data collection efforts, and consequently, influence the market information infrastructure.

Our first set of analyses examines the extent to which S&P's data collection lags are determined by the strength of the demand the data aggregator faces for corporate accounting information. We identify institutional investors as the key source of demand for the services of capital market data aggregators, and find that Standard & Poor's firm-specific data collection lag varies inversely with the firm's overall level of institutional ownership, as well as the nature of institutional investor clientele. Bushee (1998, 2001) identifies three types of institutional

[2] Consistent with the above quote, Standard & Poor's views the rigor and objectivity of its data collection process, its focus on standardization of data to increase comparability, and its emphasis on timely data dissemination as the key benefits perceived by its customers (based on Standard & Poor's Webex session titled "Compustat Data: Source to Income Statement" dated January 4, 2007, and conversations with Compustat staff members). http://www2.standardandpoors.com/portal/site/sp/en/us/page.product/dataservices_compustat/2,9,2,0,0,0,0,0,0,0,0,0,0,0,0,0,0.html (access February 2007). Consistent with the client's focus on timely availability of data, Standard & Poor's offers appropriate financial incentives to its employees in its data aggregation unit.

investors by trading behavior: transient investors (characterized by high diversification, high turnover, and shorter horizons); dedicated investors (characterized by "relation investing" in a few selected firms, with low turnover and longer horizons); and quasi-indexers (characterized by high diversification, but low turnover). Specifically, while quasi-indexers depend on corporate financial disclosures as a low-cost performance-monitoring mechanism, neither transient nor dedicated institutions rely on quick access to standardized corporate accounting information to meet their investment objectives (Bushee and Noe, 2000). Consistent with this intuition, we predict and find that the relationship between institutional ownership and dissemination lag is driven by the informational demands of quasi-indexers.

We also examine several investment characteristics that influence the demand for speedy dissemination of corporate information. Demand for corporate accounting information is likely to be lower when high arbitrage or transaction costs deter sophisticated investors from exploiting accounting-based mispricing (see, Collins et al. 2003; Ali et al. 2003; Mashruwala et al. 2006). Given the limited investment value to institutional investors, we predict and find delayed dissemination of corporate accounting information when stocks have higher arbitrage risk and transaction costs.

Due to the heightened investor awareness for index stocks (Wurgler and Zhuravskaya, 2002; Chen et al., 2004), we predict greater demand from institutional equity clients for information on stocks included in major market indices, and more so for large cap indices compared to medium or small cap indices. Results are consistent with hypothesized collection priorities, documenting a 40 percent faster collection of information from periodic SEC filings for companies in the S&P 500 index, with market capitalization and memberships in other domestic major market indices having smaller impacts. Interestingly, we find that firms that exit

a major index retain some of their prior collection priority, consistent with the evidence in Chen et al. (2004) on continued investor awareness of stocks even after they exit the S&P 500 index. We also find that membership in a major index (or the factors that lead to it) may be more important to trigger investor awareness and demand for information than merely large market capitalization. Finally, although our focus is on the institutional investor demand, we find that the dissemination lag is shorter for stocks of companies with credit ratings, consistent with credit market demand for faster dissemination.

While our hypotheses pertain to the demand side, we control for various supply-side factors likely to influence collection lags: production-capacity constraints, peak-period shifts in resources, firm-specific differences in collection efforts, and enhanced technological efficiencies. We find higher collection lags during peak times, consistent with limits to production capacity and the inability to "stock" information in anticipation of demand. We also find that the collection lag is concave in the magnitude of the backlog, consistent with the notion that S&P deploys increased resources during peak periods. Reinforcing the expectation of resource shifts during peak periods, we also find a negative association between the intensity of information arrival and the delay in information dissemination. Finally, the supply shock introduced by EDGAR has resulted in roughly a 50 percent reduction in the average collection lag. Collectively, our demand and supply factors explain a significant cross-sectional variation in S&P's data collection lag.

Our second set of analyses focuses on the impact of information dissemination speed on the trading decisions of various types of institutional investors: transient investors, quasi-indexers, and dedicated investors. Based on extant research (e.g., Bushee and Noe, 2000), we predict that transient investors are more likely to be drawn to firms whose information

environment is enhanced by faster dissemination of corporate accounting data, which facilitates the implementation of their short-run trading strategies. Since financial disclosures offer a low-cost monitoring mechanism for quasi-indexers, we also predict that quasi-indexers will be drawn to firms with faster information dissemination speeds. Dedicated investors, however, have longer horizons and invest in a few selected firms. They are less dependent on public channels of communication, and might arguably even prefer disclosure environments that enable them to retain their informational advantages (Bushee and Noe, 2000). Consequently, we make no directional predictions about the effect of dissemination speed on the trading decisions of dedicated investors.

Our long-run association analysis indicates that, after controlling for various determinants of institutional ownership (Bushee and Noe, 2000), transient institutions and quasi-indexers have larger (smaller) holdings in stocks that exhibit consistent inter-temporal increase in dissemination speed (lag) during our sample period. For instance, for firms that experienced persistent increase (decrease) in the dissemination speed over the entire sample period, the ownership of transient institutions would have increased (decreased), on average, by 1.7 (1.5) percentage points relative to firms with stable information dissemination speed.

In our short-term analysis, after controlling for various determinants of changes in institutional ownership (Ke and Petroni, 2004; Ke and Ramalingegowda, 2005; Bushee and Noe, 2000), we find that quarter-to-quarter changes in ownership by transient investors and quasi-indexers are a positive function of the speed with which corporate accounting information is disseminated by S&P. For a shift in dissemination speed equal to its inter-quartile range, the short-run effects account for around 0.32 percentage points of the quarterly change in institutional ownership. Although arguments for a directional relationship are not as persuasive

for dedicated institutions, we do find that dedicated institutions increase their ownership position when dissemination of data from periodic filings is delayed. One possibility is that these investors do prefer coarse information environments that enable them to retain their competitive informational advantage (Bushee and Noe, 2000).

Taken together, our empirical findings should be of interest to academic researchers, the investment community, and other market participants. Our study takes a first look at factors influencing the speed of dissemination of corporate accounting information by data aggregators. Controlling for supply-side constraints, we use demand factors to explain the equilibrium choice of information dissemination speed, providing new insights into the capital market's informational infrastructure. While data aggregators do not directly determine informational efficiency, our results suggest that they play a key role in responding to investor demand for corporate accounting information, collecting and distributing information in standardized machine readable form on a broad cross-section of companies, and facilitating sophisticated fundamental analysis. From a consequences perspective, we find that the data aggregators' choice of dissemination speed, in turn, influences differential trading decisions by the different categories of institutional investors. Overall, our paper provides evidence on the mutual effects that data aggregators and institutional investors have on each others' resource allocation decisions.

The remainder of the paper is organized as follows. Section 2 provides a brief description of the theory regarding the market for information goods, and discusses hypotheses development. Section 3 discusses sample selection and descriptive evidence. Section 4 presents results pertaining to the determinants of the speed of information dissemination, followed by an analysis

of the consequences of dissemination speed in section 5. Concluding remarks are provided in the final section.

2. Theory and Hypotheses Development

This section motivates our focus on information dissemination speed as a key choice variable for data aggregators. We first describe the unique characteristics of the market for corporate accounting information, and then discuss the demand and supply forces at work in the commercial market segment served by Compustat. We conclude this section by discussing our hypotheses development.

2.1 Market for Information Goods: the Case of Corporate Accounting Information

Consistent with the academic intuition on the market for information goods (Shapiro and Varian, 1999), most capital market data aggregators (e.g., Compustat, Bloomberg, and CRSP) segment the market at a minimum between commercial and educational customers, who are provided different product versions with different timeliness for which they are charged different prices. Although the financial data released by companies are non-excludable (Romer 1990), Compustat's data services are available only to paying customers. In this study we focus on the commercial segment of the market, where Compustat caters to the information needs of primarily money managers (including back testers) and also those of credit managers, management and valuation consultants, accounting firms, etc.[3]

[3] Within the commercial market, Compustat has different pricing arrangements based on the amount of assets under management. This pricing scheme is consistent with the increasing value of information as a function of assets under management. We are not aware of any usage-based fees charged by Compustat to its commercial customers, which is consistent with Sundarajan (2004) that such pricing strategy can be less than optimal when transaction costs of administering it is non trivial. Given the differential commercial price arrangements used by Compustat do not lead to different timeliness of data delivery among commercial customers, we view them as one market segment for purposes of this paper.

In the commercial market served by Compustat, neither demand response pricing nor anticipatory production of inventory is optimal or feasible. However, while information intermediaries are not capacity constrained for *reproduction* of information, they do face capacity limits on the original *production* of information. Given that corporate accounting data are provided in non-standard formats in financial statements and footnotes, Standard & Poor's Compustat hires data collectors with accounting expertise and trains these professionals in its detailed standardization and coding procedures.[4] The optimal production capacity for a data aggregator such as Compustat (proxied by the total data collection potential) will be determined by various cost-benefit considerations.

However, several institutional aspects of the corporate financial data aggregation market suggest that the customers neither demand nor are willing to pay for immediate access, which limits the choice of production capacity. First, while timeliness is important, these customers make sophisticated portfolio decisions that require access to extensive standardized financial data in machine-readable form on a broad cross-section of companies. Given that the stock market reacts more swiftly to public information (see Kothari 2001) than it would take a data aggregator to collect and standardize information using currently available technology, the resource allocation decisions for which Compustat and other similar databases are used must not depend on instantaneous access.[5] Second, as discussed below, the demand of Compustat commercial customers for faster access to accounting information varies substantially across companies.

[4] For instance, Standard & Poor's Compustat User's Guide is over 700 pages long with detailed information on what specific information is included in or excluded from each of the data items.

[5] The situation could change if the cost of data collection and standardization decreases substantially with technological developments such as XBRL. If instantaneous access can be made available for close to zero marginal cost, then the suppliers may be willing to offer it. In a related vein, we later provide evidence on how EDGAR increased the speed of information dissemination in capital markets.

Taken together, we argue that timeliness of data delivery is the key choice variable for a data aggregator of corporate accounting information. Given the nature of the information market, pricing decisions are less interesting even if data on commercial pricing arrangements were publicly available. Therefore, the focus of the paper is to examine the demand considerations that are associated with the speed with which Compustat disseminates accounting information to its commercial customers, subject to supply constraints.

2.2 Hypotheses Development

We develop hypotheses to examine the influence of institutional investors on the speed of information dissemination by S&P as well as the consequences of the dissemination speed on institutional ownership.

2.2.1 The Influence of Institutional Investors on Dissemination Speed

As discussed earlier, institutional investors (especially, money managers) are the key commercial subscribers to Compustat's corporate accounting databases, which suggests a demand for information on stocks held by institutional owners in general. In addition, our discussions with Compustat indicate that special requests by clients receive the *highest* data collection priority, and Compustat does expedite data collection efforts in such cases.[6] Such special requests are likely to further speed up information dissemination in stocks typically held by institutions.[7] Assuming that Compustat's collection efforts are geared toward meeting its customer needs, we predict the following (hypotheses stated in alternative form):

H1: There is a negative association between institutional ownership and the delay in the dissemination of corporate accounting information.

[6] See also http://www2.standardandpoors.com/spf/pdf/products/Compustat2006.pdf (accessed February 2007).
[7] We find several instances in the portfolio holdings disclosure policy section of registration statements where mutual funds indicate that they provide access to information on portfolio holdings to data aggregators such as Standard & Poor's and Vestek. Examples include the registration statements of Pacific Select Fund, Janus Capital Management, Forward Funds, The Roxbury Funds, and WT Mutual Fund. These disclosures are consistent with S&P's statement that its clients request more timely data gathering on selected companies at selected times.

Demand for information on a particular stock is likely to be influenced by differences in investment styles and portfolio objectives among institutional investors. Past literature generally groups institutional investors into three broad categories, following Bushee (1998, 2001): ***transient*** investors (characterized by high diversification, high turnover, and shorter horizons); ***quasi-indexers*** (that exhibit high diversification, but low turnover); and ***dedicated*** investors (characterized by "relation investing", with limited diversification, low turnover, and longer horizons).

Given that quasi-indexers appear to depend on corporate accounting disclosures as a low cost performance monitoring mechanism, Standard & Poor's is likely to expedite its data collection efforts for stocks favored by them. In contrast, dedicated investors do not rely on public channels of communication, and therefore, data aggregators are likely to be indifferent to stocks held by these institutional investors (Bushee and Noe 2000; Porter 1992). While the trading decisions of transient institutions could be indirectly affected by the dissemination speed (a proxy for the richness of information environment), Bushee and Noe (2000) argue that these institutions may not necessarily *rely* on the accounting information provided by information intermediaries.[8] Overall, data aggregators are more likely to focus on collecting accounting information of firms targeted by quasi-indexers. Consequently, we expect the level of holdings of quasi-indexers alone to impact dissemination speed, and hypothesize that:

> H2: There is a negative (no) association between quasi-indexer (transient/dedicated) institutional ownership and the delay in the dissemination of corporate accounting information.

[8] Our discussions with the chief operating officer and a member of his quantitative research team of an international investment management firm with client assets close to $150 billion are consistent with this intuition. Their relational investment ("fundamental analysis") portfolio staff closely monitors company financial performance (including reviewing filings, press releases, other market information, meeting with management, etc.), but do not rely on standardized panel data provided by information aggregators. On the other hand, their quantitative research portfolio (akin to quasi-indexers) staff has little contact with company management, but employs data-driven statistical analysis that relies heavily on data provided by data aggregators.

While H1 and H2 focus on the investor characteristics, we next turn to stock-specific characteristics that influence institutional investors' general demand and special requests for information. Collins et al. (2003) argue that sophisticated market participants, such as institutional investors, are unable to benefit from mispricing opportunities due to arbitrage risk and transaction costs (see Ali, Hwang et al. 2003; Mashruwala et al. 2006). We, therefore, argue that the institutional investor demand for corporate accounting information is likely to be greater when stocks have lower arbitrage risk or transaction costs, leading to the following hypothesis:

> H3a: The delay in the dissemination of corporate accounting information is lower for stocks with lower arbitrage risk or transaction costs.

Another indication of demand for a stock held by institutional investors is its membership in a major stock index (Wurgler and Zhuravskaya 2002). We expect greater demand for corporate accounting information of stocks included in major market indices, and more so, for stocks included in large cap indices compared to medium or small cap indices.[9] Our discussions with Compustat reinforce the notion of such data collection priorities, and lead us to predict the following:

> H3b: The marginal reduction in the dissemination lag of corporate accounting information is the highest for stocks included in the S&P 500 index, followed by smaller effects for stocks with market capitalization of $1 billion or more, and stocks in other S&P domestic major indices (S&P MidCap 400 and S&P SmallCap 600).[10]

2.2.2 The Consequences of Dissemination Speed on Institutional Ownership

Our last set of hypotheses focuses on the impact of dissemination speed on the trading decisions of the three types of institutional investors. Bushee and Noe (2000) find that transient

[9] See Bos and Ruotolo (2000) for the criteria for inclusion of a stock in one of the S&P domestic indices.
[10] The hypothesis reflects S&P's stated collection priorities. Consequently, one could argue that H3b is more of a test of whether Compustat's actual collection efforts match its stated guidelines rather than a direct test of an economic hypothesis. However, given that Compustat's guidelines are shaped by its customers' demand for corporate accounting information, our statistical tests should be relevant for understanding cross-sectional variations in the dissemination of corporate accounting information driven by customer needs.

institutions invest more heavily in stocks with richer information environment (proxied by higher AIMR disclosure rankings), and add to their holdings in response to increases in disclosure rankings. The key insight from Bushee and Noe (2000) is that while transient investors may not necessarily rely on the accounting information provided by information intermediaries, their trading decisions are likely influenced by the dissemination speed. In other words, we predict that transient investors are more likely to be drawn to firms whose information environment is enhanced by faster information dissemination, which would facilitate the implementation of their short-run trading strategies. Given that financial disclosures offer a low-cost monitoring mechanism for quasi-indexers, we also predict that quasi-indexers will be drawn to firms with faster information dissemination speed (Bushee and Noe, 2000).

Dedicated investors, however, have long horizons and invest in a few selected firms. They are less dependent on public channels of communication, and might arguably even prefer disclosure environments that enable them to retain their informational advantages (Bushee and Noe, 2000). Consequently, we make no directional predictions about the effect of dissemination speed on the trading decisions of dedicated investors, and restrict our hypotheses to transient and quasi-indexers as follows:

> H4a: The proportion of a firm's equity securities held by transient investors and quasi-indexers increases (decreases) in firms with rising (declining) information dissemination speed over time.
>
> H4b: Current changes in transient and quasi-indexer ownership are negatively associated with prior corporate accounting information dissemination lags.

H4a suggests that transient investors are more likely to gravitate towards stocks with increasingly richer information environment due to improved opportunities for implementing their short-run strategies. Although we expect quasi-indexers to benefit from increased dissemination speed, the ex ante expectation is that the effect may be tempered compared to

transient institutions. Note that H4a considers a general association between information dissemination lag and institutional investor holding in the long run, while H4b examines short-run changes in portfolio allocations in response to the dissemination lag.

3. Sample and Descriptive Evidence on the Speed of Information Dissemination

3.1 Sample Identification

Our sample selection procedure is summarized in Table 1, Panel A. We obtain our sample from the intersection of the quarterly Compustat Preliminary History (Prelim) database and the quarterly Compustat "As First Reported" (AFR) database. While the Prelim database includes information disclosed in earnings announcement press releases, the AFR database includes financial statement information as first reported in periodic SEC filings. Given that we examine information diffusion by capital market intermediaries, we limit our sample to publicly-traded firms in the CRSP database. Based on the above considerations, we begin with a sample of 401,707 firm quarters during the calendar years 1991 through 2004.

We obtain 10-K/10-Q filing dates from the S&P's SEC filing date database and 10KWizard. We exclude observations with missing filing dates or missing production dates for the delivery of SEC filing information. A small set of observations is dropped due to possible data coding problems. The resulting SEC filing sample is 312,775 firm quarters.

3.2 Institutional Details on Compustat Data Collection Practices and Measurement of Collection Lags

We use the AFR database production dates for the S&P's Research Insight product to examine the speed with which S&P disseminates accounting information to the marketplace. The relevant production date variable is FINALQPRD, which represents the production date when a

company's final quarterly financial data from periodic SEC filings first appeared in Compustat.[11] We calculate FL_CLAG as the number of weekdays between the SEC periodic report filing date and FINALQPRD.[12]

FINALQPRD represents the Research Insight weekly CD-ROM production date, and therefore, could understate the true speed with which information is disseminated by Compustat to its commercial clients.[13] Compustat offers its products using different delivery mechanisms with different dissemination speeds. For example, Compustat FTP and Xpressfeed rely on Internet-based delivery mechanisms that provide daily updates. Even customers receiving weekly CD-ROMs through their Research Insight subscriptions can seamlessly receive daily updates of key financial statement information. Consequently, the production date that we use may not represent the earliest point in time when data may be available to all of Compustat's commercial subscribers, giving rise to measurement error in our collection lag variables. However, Panel B of Table 1 indicates that the median FL_CLAG is 15 weekdays with the inter-quartile value ranging from eight to 23 weekdays.[14] Note that FL_CLAG extends beyond two production cycles for a majority of the firm-quarters, which suggests that the measurement error is less likely to be a concern.

In Panel B of Table 1 we also provide descriptive statistics on FL_CLAG separately for the pre- (1991-1995) and post-EDGAR (1997-2004) periods. A comparison of median values indicates that EDGAR decreased the dissemination lag by 50 percent. Interestingly, the

[11] See http://www.compustat.com/support/wi/private_shrd/dataguide/finalqprd.html.
[12] We truncate FL_CLAG values when they are negative or greater than their 99th percentile values to mitigate possible data entry error.
[13] Although our focus is on measuring the earliest availability of the Compustat data, we understand that certain commercial customers may rely on the weekly CD-ROMs for their analytic needs. In which case, the use of production dates may overstate the dissemination speed given the data may not be "available" to these customers until a few days after the production date.
[14] Our descriptive evidence is consistent with Compustat's timeliness standard of 15 business days for AFR data during peak season, with lower priority observations taking longer time. As discussed below, the measurement error in our collection lag measure upwardly biases the true collection lag.

coefficient of variation increased from 0.66 to 0.90 after EDGAR, which suggests a large cross-sectional variation in dissemination speed even in the electronic filing era.

3.3 Compustat's Collection Efforts by Calendar Time

In Figure 1A we plot the average number of SEC filings (NFL) and "final updates" (NFINAL) (Compustat parlance for collection of data from 10-K/10-Q filings). The four large spikes in the arrival of periodic SEC filings correspond to the mandatory filing dates of calendar-year registrants in the pre-accelerated filing era (see Griffin 2003). In contrast, Compustat's collection efforts are distributed relatively evenly in calendar time. In Figure 1B we plot the average cumulative proportion of SEC filings and their collection over the 52 production weeks.[15] The cumulative proportion of collection efforts is closer to a diagonal straight line, representing smoother collection efforts in calendar time when compared to the release of SEC periodic reports. More importantly, the dissemination lag for periodic filings points to the presence of capital market participants who value access, albeit delayed, to standardized electronic accounting databases that provide comprehensive information on a broad cross-section of publicly-traded companies

4. Determinants of the Speed of Information Dissemination: Model and Empirical Findings

4.1 Model Development

To test our hypotheses on the determinants of information dissemination speed, we estimate the following panel-data regression:[16]

[15] Given that we graph the time-series of SEC filings and their collection *averaged by weeks*, the line for the cumulative collection in the early weeks could be above the cumulative filings. This merely reflects the fact that at the beginning of each calendar year S&P clears the backlog of filing data cumulated from the previous year while facing very few inflow of new releases.

[16] While the focus of the paper is on the dissemination of information in periodic SEC reports, we also examine the determinants of the dissemination speed of earnings press releases. Untabulated analysis show that the dissemination speed of earnings information has very limited cross-sectional variation, and the regression results show a pseudo R-squared of less than four percent. Given the importance of earnings announcements to the marketplace, the evidence

$$FL\_CLAG_{iq,tp} = \beta_0 + \beta_1 INST_{iq-1} + \beta_2 SPECTRUM_{iq-1} + \beta_{3a} ARBRISK_{iq} + \beta_{3b} PRICE_{iq} + \beta_{3c} VOLUME_{iq}$$
$$+ \beta_{4a} S\&P500_{iq} + \beta_{4b} CAP1B_{iq} + \beta_{4c} S\&P1000_{iq} + \delta_5 CREDITRATE_{iq} + \delta_{6a} FINSERV_i + \delta_{6b} UTILITY_i$$
$$+ \delta_{7a} SPECIAL_{iq} + \delta_{7b} EXTRAORD_{iq} + \delta_8 CONCUR_{iq} + \delta_{9a} EA\_BACKLOG_{t-1} + \delta_{9b} EA\_BACKLOG_{t-1}^2$$
$$+ \delta_{10a} FL\_BACKLOG_{t-1} + \delta_{10b} FL\_BACKLOG_{t-1}^2 + \delta_{11a} NEA_t + \delta_{11b} NEA_t^2 + \delta_{12a} NFL_t + \delta_{12b} NFL_t^2$$
$$+ \delta_{13} YEAREND_p + \delta_{14} QTREND_p + \delta_{15} MONDAY_p + \delta_{16} TUESDAY_p + \delta_{17} THURSDAY_p + \delta_{18} FRIDAY_p$$
$$+ \sum_{T=1992}^{2004} \beta_{19,T} YT_{iq} + \varepsilon_{iq,tp} \qquad (1)$$

where

$FL\_CLAG_{iq,\,tp}$ is the number of weekdays from the10-K/10-Q filing date t of firm i for fiscal period q to Compustat production date p (FINALQPRD) when the firm's finalized financial data first appeared in Compustat's Research Insight database;

$INST_{iq-1}$ is a proxy for institutional ownership. In the first version of the model, institutional ownership is captured by $INSTOWN_{iq-1}$, the percentage of shares of firm i held by institutional investors as of the beginning of calendar period q (H1). In the second version of the model, $INSTOWN_{iq-1}$ is replaced by $TRAOWN_{iq-1}$, $QIXOWN_{iq-1,}$ and $DEDOWN_{iq-1}$, capturing the percentage of ownership at the beginning of calendar quarter q by transient, quasi-indexer and dedicated institutions, respectively (H2);[17]

$ARBRISK_{iq}$ is the standard deviation of market model residuals estimated over the fiscal period q, multiplied by 100, winsorized at 99.5 percentile (H3a);

$PRICE_{iq}$ is the closing stock price of firm i as of the end of fiscal period q, winsorized at 99.5 percentile (H3a);

$VOLUME_{iq}$ is the daily trading volume in millions of dollars of firm i, averaged over the fiscal period q, winsorized at 99.5 percentile (H3a);

$S\&P500_{iq}$ takes the value of one if firm i is included in the S&P 500 index in fiscal period q, and zero otherwise (H3b);

---

indicates that S&P shifts its resources and collects earnings announcements within three business days in most cases, resulting in limited variation that can be explained by the posited determinants. However, many of the determinants of earnings announcement collection lags are statistically significant in the predicted directions. In the interest of parsimony, we restrict our discussion to the SEC filing regression results.

[17] We obtain institutional ownership data from CDA Spectrum which extracts the data from the SEC's Form 13(f). Form 13(f) must be filed each calendar quarter by all institutions with greater than $100 million in equity securities, and therefore, the institutional ownership data is available only on a calendar-quarter basis. Consequently, we use the time subscript "q-1" to denote that we use the institutional ownership of the calendar quarter immediately preceding the fiscal quarter q.

$CAP1B_{iq}$ takes the value of one if firm i's market value of equity as of the end of fiscal quarter q is at least $1 billion but the firm is not included in the S&P 500 index, and zero otherwise (H3b); and

$S\&P1000_{iq}$ takes the value of one if firm i is included either in the S&P MidCap 400 index or the S&P SmallCap 600 index in fiscal period q but has market capitalization less than $1 billion, and zero otherwise (H3b).

Control Variables

$FINSERV_i$ equals one if firm i is a financial services firm (SIC 6000-6999), and zero otherwise,

$UTILITY_i$ equals one if firm i is a utility firm (SIC 4900-4999), and zero otherwise,

$CREDITRATE_{iq}$ equals one if a S&P credit rating is available for firm i at the end of fiscal period q, and zero otherwise,

$EA\_BACKLOG_{t-1}$ is the number of uncollected earnings announcements (in '000s) made prior to date t by all companies followed by Compustat,

$FL\_BACKLOG_{t-1}$ is the number of uncollected 10-K/10-Q filings (in '000s) made prior to date t by all companies followed by Compustat,

NEA_t is the number of earnings announcements (in '000s) made at date t by all companies followed by Compustat,

NFL_t is the number of 10-K/10-Q filings (in '000s) made at date t by all companies followed by Compustat,

$SPECIAL_{iq}$ equals one if firm i reports any special item in SEC 10-K/10-Q filings for fiscal quarter q, and 0 otherwise,

$EXTRAORD_{iq}$ equals one if firm i reports any extraordinary item in SEC 10-K/10-Q filings for fiscal quarter q, and 0 otherwise,

$CONCUR_{iq}$ equals one if firm i's earnings release for fiscal period q concurs with its periodic SEC filing date for fiscal period q,

$YEAREND_p$/$QTREND_p$ equals one if the production date p is the last working day of a year /quarter,

$MONDAY_p$, $TUESDAY_p$, $THURSDAY_p$ and $FRIDAY_p$ are weekday dummies for the production date p.

As in Mashruwala et al. (2006), we use stock price (PRICE) and dollar trading volume (VOLUME) as an inverse measure of transaction costs. Consistent with Mashruwala et al. (2006), Pontiff (1996) and Wurgler and Zhuravskaya (2002), we use idiosyncratic volatility (ARBRISK) as a proxy for arbitrage risk.

We include several control variables in our model, in addition to proxies for the main variables of interest. Additional demand-side control variables include membership in a regulated industry (banks, insurance companies and utilities), and the existence of S&P credit ratings. We consider the possibility that the demand for timely dissemination of accounting information may be limited for firms in regulated industries for at least two reasons. First, the standardized data collection format used in the standard Compustat database may be less suitable for analyzing financial services firms, so commercial customers may rely on other readily available data sources that better reflect industry-specific idiosyncrasies (e.g., FERC filings, NAIC releases, Call Reports, SNL Datasource, and Bank Compustat Database). Second, prior research (Teets 1992; Teets and Wasley 1996) indicates that regulated utilities have substantially lower earnings response coefficients compared to non-regulated firms, which suggests that there may be less demand for accounting information of utilities. We control for firms with S& P credit ratings because S&P's credit rating services rely on Compustat databases as part of its ratings determination process (private correspondence), and analysts and managers involved in credit evaluations constitute an important group of external customers as well.

We also control for supply factors that capture Compustat's capacity constraints as well as its resource allocation decisions in response to peak and non-peak demands. We use current backlog of uncollected corporate accounting information to proxy for the effect of capacity constraints, and incorporate the squared value of current backlog to capture increased Compustat

resource allocation during peak times.[18] To examine any crossover effects of earnings announcement backlog on the dissemination speed of information in SEC filings, we include separate backlog variables for earning announcement and SEC filings.

Intensity of information arrival is likely to influence dissemination speed, so we also control for the volume of earnings announcement and SEC filing information arrival. Similar to our backlog variables, we include both linear and quadratic terms for information arrival intensity as well as consider any crossover effects between earnings announcements and SEC filings.

Finally, we control for days of the week, fiscal period end, and instances where earnings announcements and filing dates coincide. We include controls for the incidence of special or extraordinary items, as firms reporting these items are likely to experience business events that influence data aggregation complexity, and consequently affect dissemination speed. We include calendar-year dummies to examine the supply side effect of EDGAR on information dissemination.[19]

4.2 Descriptive Statistics

We provide descriptive statistics on firm-specific variables in Panel A of Table 2, and on macro-level control variables in Panel B of Table 2. Descriptive evidence on institutional ownership and the proxies for arbitrage risk and transaction costs in Panel A is consistent with prior research (e.g., Collins et al. 2003; Mashruwala et al. 2006). Approximately 20 percent of the sample firm-quarters have an S&P credit rating, and slightly over 25 percent of the sample

[18] For example, some experienced Compustat data collectors do get promoted and are given professional opportunities outside its data collections group. However, Compustat finds it expedient to leverage the data collection expertise of these individuals during peak time periods. The institutional knowledge and expertise of these individual continue to be highly relevant for Compustat's data collection efforts.

[19] Although EDGAR became effective on May 6, 1996 (Balsam et al. 2002), our discussions with S&P indicate that during 1996 Compustat was transitioning from the paper version of the filings to EDGAR, and therefore, may not have obtained the full benefits of EDGAR.

observations are included in Compustat's priority categories based on membership in major market indices or firm size. Regulated firms comprise 22 percent of the firm-quarters, and slightly over 26 percent of sample firms report special or extraordinary items in their 10-Q/10-K filings. The incidence of firms providing earnings information in the periodic SEC filings without a preliminary earnings announcement or providing an earnings press release concurrently with the periodic filing is around 20 percent (see Amir and Livnat 2005).

Panel B of Table 2 indicates that the standard deviations of NPRELIM and NEA are roughly comparable (0.384 versus 0.385), which suggests that the time-series volatility in Compustat's collections efforts for earnings announcements is similar to that of the arrival of earnings announcements. In contrast, the standard deviation of NFL (0.766) is more than double that of NFINAL (0.359), indicating a smoother inter-temporal collection effort for filings, and suggesting a higher priority given to the collection of earnings announcements.

4.3 Regression Results

Table 3 shows the results of estimating model (1) using Poisson regression. We present results for two versions of the model with the institutional ownership broken down into three categories in the second version. Except when examining the institutional ownership effects, we limit our discussions to the first version of the regression. We report Huber-White standard errors which are adjusted for heteroskedasticity and firm-specific clustering in our panel data (see Rogers 1993; Williams 2000; Petersen 2007). We also report the marginal effects as the expected percentage change in the collection lag of SEC filings at the inter-quartile range for all continuous variables and for a unit change for all dummy variables.

The test and control variables explain a non-trivial variation in the dissemination speed as indicated by a 23 percent pseudo R-squared. In addition, the results indicate that both hypotheses

1 and 2 are strongly supported. Institutional ownership (proxy for general demand and customers' special requests) is statistically significant in explaining Compustat collection efforts, with a marginal effect of 5.19 percent decline in dissemination lag. Moreover, when total institutional ownership is decomposed into the three categories of institutional investors discussed earlier, the result for institutional ownership in our first regression is driven by the significantly negative association between quasi-indexer ownership and accounting information collection lag. As predicted in H2, we find that Compustat collects data faster for firms with larger quasi-indexer ownership, consistent with these investors placing a premium on speedy access to accounting information in a comparable format for ease of performance monitoring and implementing their portfolio strategy. Neither transient nor dedicated investor ownership is significantly associated with information dissemination speed, suggesting that these investors do not rely on accounting information provided by data aggregators as the primary source of information for their investment decisions.

Results also support the notion that stock-specific characteristics influence institutional demand for accounting information. The coefficients on PRICE and VOLUME are significantly negative, and on ARBRISK significantly positive, suggesting that demand for corporate accounting information is greater (and consequently collection speed is faster) when stocks have lower transaction costs or arbitrage risk, consistent with hypothesis 3a.

Evidence strongly suggests that institutional investors demand faster dissemination of information on stocks included in major market indices or large market capitalization, with membership in the S&P 500 index resulting in the largest (40 percent) reduction in the dissemination lag as expected. We find a 22 percent reduction in collection lag for stocks in the S&P MidCap 400 and S&P SmallCap 600 indices. The collection lag for large firms (greater

than $1 billion market capitalization) is also economically significant (13 percent), but contrary to the collection hierarchy posited in H3b, the dissemination lag is shorter for firms in mid and small cap indices than for large firms.[20]

In the interest of brevity, we only discuss results relating to selected control variables, although most are associated with collection lag in the expected direction. Consistent with dissemination speed being influenced both by external demand from credit managers and internal demand from S&P credit rating service, we find that collection lags for firms with a S&P credit rating are significantly lower (around four percent). We also find a significant delay in the dissemination of standardized corporate accounting information of regulated firms, possibly because of industry-specific information idiosyncrasies discussed earlier, and/or lower demand for information on regulated firms. The effect is more salient for utilities, where there is a 23 percent delay in the dissemination of information in periodic SEC filings.

Although the focus of our analysis is on the demand side, our regression provides several insights on the supply-side effects as well. Compustat's collection lag increases by 6.9 percent for firms facing economic circumstances that lead to extraordinary accounting gains or losses, consistent with increased efforts for data aggregators in collecting accounting information. The slope coefficients of the linear and quadratic terms of the filing backlog variables have the correct sign and are statistically significant. The collection lag increases by over 27 percent when the backlog of uncollected filing data increases by its inter-quartile range, with the increased peak-time collection efforts bringing down the lag by roughly six percent. The slopes for the backlog variable "FILEBACKLOG" (0.18) and the proxy for the intensity of SEC filings "NFILE" (0.15) are roughly of the same magnitude, suggesting that Compustat's resource

[20] As a sensitivity check, we also include three independent variables interacting quasi-indexer ownership with the three dummies for membership in market indices and large firms. The slope estimates for the interaction terms exhibit monotonicity consistent with the dissemination hierarchy discussed in H3b.

allocations for collecting periodic SEC filings are not further heightened by the increased intensity of information arrival during peak periods.[21] Finally, our evidence shows a dramatic improvement in the dissemination speed of periodic filing information following the implementation of EDGAR, indicated by a large drop in its collection lag beginning 1997.[22]

To provide some intuition behind the relative explanatory power of the various determinants, we regress FILE_CLAG separately on the demand-side variables, supply-side variables, and calendar-year dummies. The pseudo R-squared for the supply-side variables (12.4 percent) is more than twice that of the demand-side variables (5.7 percent). The larger supply-side effects suggest that S&P's production choices could have nontrivial effects on institutional investor trading (see Section 5). The pseudo R-squared for time dummies is the largest (15.3 percent), which indicates the large positive externality of EDGAR on S&P's collection lag.

4.4 Effect of index switches on dissemination speed

Chen et al. (2004) find that the inclusion in S&P 500 leads to an increase in the investor awareness of a stock, which does not vanish if the stock is subsequently deleted from the index. To the extent changes in investor awareness is asymmetric to additions and deletions from major market indices, one would expect a corresponding asymmetric change in dissemination lag. To examine this, we expand the first version of the regression model in Table 3 to incorporate dummy variables indicating switches among "index" memberships (i.e., two index groups and one firm-size group) and examine their marginal effect on information dissemination speed. For

[21] Note that backlog at the beginning of a period plus intensity of information arrival during the period minus collections during the period equals backlog at the end of the period.

[22] As a sensitivity check, we use a measure of relative dissemination speed that is based on a Z-transformation of FILE_CLAG of all firms in a given calendar quarter (Blom, 1958). The tenor of our OLS regression results using the Z-transformed FILE_CLAG remains unchanged except that trading volume is no longer significant. As one would expect, we do not find any significance for the EDGAR effect as the focus is on the relative speed.

parsimony, we present in Table 4 the parameter estimates of only the dummy variables relating to index memberships and changes. The inferences from other variables remain unchanged.

The coefficients on the diagonal show the effect on dissemination lag from maintaining membership in a particular index throughout our sample period. (e.g., membership in the S&P 500 index is associated with a marginal reduction of 39.61% on dissemination lag). The coefficients above (below) the diagonal represent the incremental effect on dissemination lag of entry into an index with higher (lower) priority with a corresponding exit from an index with a lower (higher) priority. The diagonal elements indicate that once we control for switches among indices, we do find a monotonic increase in the dissemination speed as we move from the S&P 1000 to S&P 500 priority groups, consistent with S&P's stated priority.

The behavior of the off-diagonal elements not only is consistent with the investor awareness argument in Chen et al. (2004), but also provides evidence on how investor awareness may impact Standard & Poor's stated priority. When firms are first added to the lowest priority index (S&P1000), S&P's collection lag decreases by 12 percent (two-tailed p-value = 0.108), with an additional decline of 15 percent (p-value < 0.001) when their market capitalization goes above $1 billion, and another 5.8 percent reduction when they move to S&P 500 (p-value = 0.063). In terms of downward movements in priority, even when an S&P 500 firm ceases to have an "index" membership, its dissemination lag continues to be 12 percent shorter than other firms without an "index" membership (Chen et al., 2004). Consistent with Shankar and Miller (2006) Standard & Poor's does not retain any dissemination priority for stocks that were hitherto part of S&P 1000.

Note, however, that when the market capitalization of a stock that was not in a major index increases above $1 billion, there is no statistically significant change in the dissemination

lag. In contrast, when a CAP1B stock enters S&P 1000 its dissemination lag decreases by 21 percent. This suggests that membership in a major index (or the factors that lead to it) may be more important to trigger investor awareness and demand for information than large market capitalization. Taken together, while H3b is based on Standard & Poor's *stated* collection priority, our analysis suggests that its customers' *actual* demand leads to a faster delivery of corporate accounting information of stocks in major indices.

5. Consequences of Dissemination Speed on Institutional Ownership: Model and Empirical Findings

In Section 5.1 we provide descriptive evidence on the inter-temporal characteristics of firms that experience increasing or decreasing dissemination speed over our sample period. The results of our tests of H4a and H4b are discussed in the following two sub-sections.

5.1 Descriptive Evidence on Inter-Temporal Changes in Dissemination Speed and Firm Characteristics

We first identity firms whose information dissemination *speed* has consistently increased over the 1991-2004 timeframe versus those for which the dissemination *lag* has increased during the same period. To insulate our analysis from the inter-temporal effects of EDGAR, we conduct a Z-transformation of FILE_CLAG of all firms in a given calendar quarter following Blom (1958). Then for each firm with at least eight quarters data, we regress the Z-transformed FILE_CLAG on a time trend variable, and assign firms into the SPEED (SLOW) group if the coefficient on the time trend is significantly negative (positive) at a two-tailed p-value of 0.05. Unassigned firms are considered as having a stable inter-temporal dissemination speed. Overall, of the 10,279 firms in our panel data, we identify 1,319 (674) SPEED (SLOW) firms.

Figure 2 compares the dissemination speed and other firm characteristics between firms whose collection lags have trended upwards versus downwards over time. All values are reported as deviations from those of the "stable" firms.[23] The directional evidence in Figure 2A shows, by construction, two sets of firms with divergent dissemination speeds over the sample period. In terms of magnitude, the evidence shows an economically significant increase (decline) of around 10 (20) days in collection lag of the SPEED (SLOW) group. Except for the S&P 1000 membership variable, the time trend of all other firm characteristics shown in Figures 2B through 2H is consistent with information environment differences between the two groups. For instance, firms in the SPEED group have gained an additional 15 percent institutional ownership compared to the SLOW group. Similarly, firms in the SPEED group have mustered roughly three additional equity analysts relative to the SLOW group.

5.2 Association Between Dissemination Lag and Institutional Ownership

To test H4a, we estimate model (2) to examine the association between dissemination speed and ownership of the stock by different categories of institutional investors:

$$\begin{aligned} INST_{iq} &= \beta_0 + \beta_1 TREND_{iq} + \beta_2 TREND_{iq} \times SPEED_i + \beta_3 TREND_{iq} \times SLOW_i + \delta_4 MRET_{iq} \\ &+ \delta_5 TVOL_{iq} + \delta_6 Ln(MVE_{iq}) + \delta_7 BETA_{iq} + \delta_8 IRISK_{iq} + \delta_9 LEV_{iq} + \delta_{10} DP_{iq} + \delta_{11} EP_{iq-1} \\ &+ \delta_{12} BP_{iq} + \delta_{13} SGR_{iq} + \delta_{14} RATE_{iq} + \delta_{15} S\&P500_{iq} + \varepsilon_{iq} \end{aligned} \quad (2)$$

We include TREND, which sequentially takes a value of 1 through 55 for the quarters 1991Q2 to 2004Q4, to control for the documented time trend in institutional ownership (Gompers and Metrick, 1998; Bushee and Noe, 2000). We set the dummy variable SPEED (SLOW) equal to 1 if the firm belongs to the SPEED (SLOW) group. To test H4a, we interact SPEED (SLOW) with TREND in our panel data regression with an expectation of a significantly positive (negative) coefficient on the interaction term for transient institutions and quasi-

[23] Although we use Z-transformed variable to identify speeding and slowing firms, we plot the untransformed values for descriptive purposes.

indexers. In addition, we include a set of variables from Bushee and Noe (2000) to control for other factors shown to be associated with institutional ownership (defined in Table 5).

Table 5, Panel A presents the results of estimating model (2). As discussed in Peterson (2007), we use standard errors corrected for both clustering by firm and time. The coefficient of TREND is significantly positive in the regressions with TRAOWN and QIXOWN as dependent variables, i.e., on average, the percentage of equity ownership of transient investors as well as quasi-indexers has grown over time. In contrast, the ownership by dedicated institutions has declined during the same period.

After we control for the macro trend in institutional ownership and other firm characteristics, the results in Table 5 are largely consistent with hypothesis 4a. The coefficient of TREND_SPEED is significantly positive for both transient institutions and quasi-indexers, which suggests that these investors gravitate towards stocks with a finer information environment. The coefficient of TREND_SLOW is in the predicted direction for both transient investors and quasi-indexers, but statistically significant only for the transient investors. We make no specific predictions relating to dedicated investors, and find no significant results for either of the two interaction terms relating to dedicated institutions.[24] In sum, not only are transient investors drawn to firms with richer information environment, but also they shun stocks with slower information dissemination speed. However, while quasi-indexer ownership of a stock increases with the dissemination speed of its accounting information, it does not show a statistically significant relation when the speed declines.

In summary, while various other firm characteristics may have a larger effect, the economic significance of dissemination speed is far from trivial. For instance, the slope estimates

[24] We do not discuss the results for the control variables as they are mostly consistent with the evidence in Bushee and Noe (2000). However, some of the differences in the results (e.g., the slope estimates of S&P500) could be explained by their focus on firms with AIMR ratings as opposed to a broader set of firms in our sample.

for transient investors suggest that for firms that experienced persistent increase (decrease) in the dissemination speed over the entire sample period, the ownership of transient institutions would have increased (decreased), on average, by 1.7 (1.5) percentage points.

5.3 Institutional Investor Response to Changes in Information Environment

Our previous analysis provided evidence of long-run association between dissemination speed and institutional ownership. We next examine the short-run response of institutional investors to dissemination speed. To test H4b, we follow Ke and Petroni (2004) (see also Ke and Ramalingegowda, 2005) and estimate the following model to test the influence of dissemination speed on institutional trading behaviors:

$$\Delta INST_{iq} = \beta_0 + \beta_1 FL\_CLAG\_KNOWN_{iq-1} + \beta_2 UNKNOWN_{iq-1} + \delta_3 INST_{iq-1} + \delta_4 PORTWT_{iq-1} + \delta_5 Ln(MVE_{iq-1}) + \delta_6 BTM_{iq-1} + \delta_7 RETQ24_{iq} + \delta_8 RETQ1_{iq} + \delta_9 RETQ0_{iq} + \delta_{10} UE_{iq} + \varepsilon_{iq} \quad (3)$$

where

$INST_{iq}$ represents $TRAOWN_{iq}$, $QIXOWN_{iq}$ and $DEDOWN_{iq}$ measured as the percentage of share ownership at the end of a calendar quarter q of firm i by transient, quasi-indexing, and dedicated institutions, the subscript q-1 represents ownership at the beginning of the quarter, and Δ denotes corresponding changes over a calendar quarter. $PORTWT_{iq-1}$ is the mean portfolio weight of shares of firm i in quarter q in the holdings of the institutions (weighted by the total market cap of each institution's stock portfolio at the beginning of quarter q).

$UNKNOWN_{iq-1}$ equals to 1 if by the end of quarter q the firm have not yet filed their periodic SEC report for the prior quarter or Standard & Poor's has not yet collected the information from the periodic report, and 0 otherwise. FL_CLAG_KNOWN equals FL_CLAG (zero) when UNKNOWN equals zero (one).

$Ln(MVE)_{iq-1}$ is the natural logarithm of market value of common equity at the beginning of fiscal quarter q. BTM_{iq-1} is the ratio of book value of equity to market value of equity at the beginning of fiscal quarter q. $RETQ24_{iq}$ ($RETQ1_{iq}$) is the buy-and-hold raw return for two through four calendar quarter (one calendar quarter) prior to the beginning of calendar quarter q. $RETQ0_{iq}$ is the buy-and-hold raw return from 30 to 3 trading days before the earnings announcement date of quarter q. UE_{iq} is change in earnings before extraordinary items from fiscal quarter q-4 to quarter q.

The dependent variable in model (3) is the quarter-to-quarter change in the institutional ownership of a given firm. In terms of the hypothesized effect, we consider the actual dissemination lag (FL_CLAG) as the variable of interest. Given that institutional investors' trading behavior would be conditional on their expectations regarding the information environment, we need a suitable measure of unexpected changes in dissemination speed to test H4b. Unreported firm-specific analysis suggests that, on average, FL_CLAG does not exhibit significant inter-temporal persistence. When we regress FL_CLAG on its one-period lagged values for each firm, we find that the median slope is 0.070 with 50 percent of observations falling between -0.095 and 0.254.[25] Given the lack of an inter-temporal persistence in FL_CLAG, we consider the *level* of the lagged dissemination speed as the predictor of changes in institutional ownership.

Note, however, that for certain firms the actual dissemination of corporate accounting information may not have occurred by the end of the current calendar quarter (q) because either the firms have not yet filed their periodic SEC reports or Standard & Poor's has not yet collected the information. We include a dummy variable (UNKNOWN) for these cases. We then define the collection lag variable "FL_CLAG_KNOWN" as equal to the actual collection lag "FL_CLAG" (zero) when UNKNOWN equals zero (one). Based on H4b we predict a negative slope coefficient for FL_ CLAG_KNOWN in the models with changes in transient institution and quasi-indexer ownership as the dependent variable. With respect to UNKNOWN, while we do not make specific predictions, the slope would have the same sign to the extent UNKNOWN proxies for the coarseness of the information environment. In addition to the variables of interest, we consider several control variables, as identified in Ke and Petroni (2004), and include fixed

[25] To control for the EDGAR effect, we subtract the cross-sectional mean prior to estimating the AR(1) parameter. The slope estimates are very similar when the regression is based on the Z-transformed FL_CLAG.

effects for firms and for time periods. By including fixed effects, we effectively exclude any firm and time effects from our collection lag variable.

The results of estimating model (3), with standard errors corrected for clustering by both firm and time, are reported in Table 6. Consistent with H4b, the coefficient of FL_CLAG_KNOWN is significantly negative in the ΔTRAOWN and ΔQIXOWN models. For a shift in dissemination speed equal to its inter-quartile range (15 weekdays) the short-run effects account for around 0.32 (0.20) percentage points of the quarterly change in transient (quasi-indexer) institutional ownership. While we do not make specific predictions for dedicated institutions, we do find that dedicated investors in fact increase their ownership interest in response to slowing information dissemination. As suggested by Bushee and Noe (2000), a possible explanation is that dedicated investors, who have private channels of communication with investee firms, might prefer disclosure environments that facilitate the retention of their informational advantages, and therefore, increase their portfolio weights of such stocks. Interestingly, in all three regressions, the coefficient of UNKNOWN has the same sign as that of the collection lag variable, and is statistically significant.[26]

The slope estimates for the control variables are largely consistent with expectations. The statistically significant negative coefficients for OWN_{iq-1} and $PORTWT_{iq-1}$ indicate that transient and quasi-indexer institutions periodically monitor their positions to maintain a well-diversified portfolio. Dedicated institutions are less sensitive to past portfolio weight in a specific firm due to their relational investing style that relies less on diversification. The strong positive effect of firm size on institutional ownership is consistent with transient investors (quasi-indexers) being drawn to stocks with finer information environment that leads to lower price impact of trades (a

[26] Given that UNKNOWN is more likely to take the value of one during the fourth fiscal quarter than during the interim quarters, we modify model (3) and estimate separate slope coefficients for UNKNOWN for interim and fourth quarters. The estimates and the statistical significance for the two coefficients are very similar.

quick access to publicly available information). Not surprisingly, only transient institutions react to short-run price changes and rebalance their portfolio. Similarly, only dedicated institutions show no sensitivity to short-term earnings information.[27]

In summary, our findings suggest that institutional investors react to variations in the speed of dissemination of corporate accounting information. However, a few observations are in order. First, while we have addressed the correlated omitted effects problem with an extensive set of controls, we cannot eliminate the possibility that dissemination speed is merely proxying for other firm characteristics that lead to changes in institutional ownership. Second, establishing causality in archival research is admittedly challenging. However, the differential predictions for the determinants of dissemination speed versus the institutional investor response provide added comfort to our results. Note that, on the determinants side, we predict and find that only quasi-indexers influence the dissemination speed. On the investor reaction side, however, we expect and find that both transient and quasi-indexer institutions respond to information dissemination speed. In addition, the opposite result for dedicated institutions, although not part of our hypotheses, is consistent with the intuition in the extant literature (Bushee and Noe, 2000). Overall, our findings indicate that not only do the data aggregators shape the information environment of firms, their dissemination choices also affect investor trading behavior.

6. Conclusion

Despite the fact that technological advances now make it possible for investors to obtain almost immediate access to accounting information released by firms in their periodic filings, the

[27] As a robustness check, we include additional control variables from Bushee and Noe (2000, Table 4) that are not considered in model (3) (i.e., lagged values of trading volume, beta, and idiosyncratic risk, and changes in leverage, dividend yield, earnings yield, sales growth, credit rating, and shares outstanding). The tenor of our results is unaffected by the inclusion of the additional control variables. Similar in spirit to Bushee and Noe (2000), we also include lagged quintiles of FILE_CLAG_KNOWN and lagged UNKNOWN and find that neither is statistically significant except for FILE_CLAG_KNOWN in the transient institutions model (negative and statistically significant).

multi-billion dollar market for the services of data aggregators has only been growing in recent years. EDGAR has had a dramatic impact on the efficiency of such data aggregators: we document a 50 percent decline in S&P's collection lag for periodic filings in the post-EDGAR period. Despite increased efficiencies over time, substantial cross-sectional variation still persists in information dissemination speed across firms. Changing technologies (e.g., XBRL) will likely further improve the overall efficiency of data aggregators. However, unless supply costs are virtually eliminated, cross-sectional differences in information dissemination speed are likely to persist, much as they have in the wake of EDGAR.

In broad terms, we hypothesize that information dissemination policies of data aggregators are driven by differences in the value that market participants place on speedy access to comparable accounting information pertaining to different firms. Information dissemination policies, in turn, influence the information environment of firms, thereby affecting the trading decisions of the investor clientele.

Our findings are consistent with these predictions. We find that S&P disseminates accounting information faster for firms with higher quasi-indexer institutional ownership, lower transaction costs (proxied by price and trading volume) and lower idiosyncratic risk. Our results are consistent with arguments in past research that quasi-indexers rely on corporate accounting information as a low-cost performance monitoring device, and that high arbitrage risk or transaction costs deter institutional investors from exploiting accounting-based mispricing. We also find that dissemination speed increases for stocks listed in major indices, possibly due to heightened investor awareness of index stocks.

On the consequences front, we find evidence that both quasi-indexers and transient investors gravitate to stocks with faster information dissemination speed. Past research has

shown that transient investors are better able to exploit their informational advantages and implement their trading strategies in the presence of a richer information environment. Quasi-indexers rely on accounting information to monitor the performance of firms in their portfolio. Our results are consistent with these arguments.

Taken together, our results provide evidence on the role played by capital market data aggregators in the informational efficiency of the marketplace. The following remark made by Merton (1987) on tests of weak form efficiency is applicable today to corporate accounting information and to our understanding of semi-strong market efficiency:

> "It is, for example, common in tests of the weak form of the Efficient Market Hypothesis to assume that real-world investors at the time of their portfolio decisions had access to the complete history of all stock returns. When, however, investors' decisions were made, the price data may not have been in reasonably-accessible form…"

While we do not suggest that data aggregators directly determine capital market informational efficiency, we argue that they act as a mechanism that responds to the needs of marketplace by disseminating corporate accounting information in standardized, comparable form at differential speeds. Their response, in turns, helps shape the information environment of firms, and affects investor trading behavior. In this respect, our study provides new evidence on the informational infrastructure of the capital market. Future research could explore other consequences of information dissemination by data aggregators, including its effects on market efficiency.

References:

Ali, A., L. Hwang, and M.A. Trombley, 2003. Arbitrage Risk and the Book-to-Market Anomaly. *Journal of Financial Economics* 69: 355-373.

Amir, E., and J. Livnat, 2005. The economic consequences of (not) issuing preliminary earnings announcement. Working paper, London Business School, New York University.

Balsam, S. E. Bartov, and C. Marquardt, 2002. Accruals Management, Investor Sophistication, and Equity Valuation: Evidence from 10-Q Filings. *Journal of Accounting Research* 40(4): 987-1012.

Blom, G., 1958. Statistical Estimates and Transformed Beta Variables, New York: John Wiley & Sons, Inc.

Bos, R. J. and M. Ruotolo, 2000, General Criteria for S&P US Index Membership. Standard & Poor's, mimeo.

Botosan, C., 1997. Disclosure Level and the Cost of Equity Capital. *The Accounting Review* 72(3): 323-349.

Bushee, B.J., 1998. The Influence of Institutional Investors on Myopic R&D Investment Behavior. *The Accounting Review* 73(#): 305-333.

Bushee, B.J. and C.F. Noe, 2000. Corporate Disclosure Practices, Institutional Investors, and Stock Return Volatility. *Journal of Accounting Research* 38(Supplement): 171-202.

Bushee, B.J., 2001. Do Institutional Investors Prefer Near-Term Earnigns over Long-Run Value? *Contemporary Accounting Research* 18(2): 207-246.

Bushee, B.J., J.E. Core, W.R. Guay, and J. Wee, 2007. The Role of the Business Press as an Information Intermediary. Working paper, University of Pennsylvania.

Chen., H., G. Noronha, and V. Singal, 2004. The Price Response to S&P 500 Index Additions and Deletions: Evidence of Asymmetry and a New Explanation. The Journal of Finance 59(4): 1901-1929.

Collins, D., G. Gong, and P. Hribar, 2003. Investor Sophistication and the Mispricing of Accruals. *Review of Accounting Studies* 8, 251-276.

Fox, J., 1997. Learn to Play the Earnings Game (and Wall Street Will Love ou. *Fortune* (March 31): 76-79.

Gompers, P.A. and A. Metrick, 1998. Institutional Investors and Equity Prices, NBER Working Papers 6723, National Bureau of Economic Research, Inc.

Griffin, P.A., 2003. Got information? Investor Response to Form 10-K and Form 10-Q EDGAR Filings. *Review of Accounting Studies* 8 (December): 433-460.

Hansen, M.T. and M. R. Haas, 2001. Competing for Attention in Knowledge Markets: Electronic Document Dissemination in a Management Consulting Company. *Administrative Science Quarterly* 46: 1-28.

Healy, P.M., A. Hutton, and K. Palepu, 1999. Stock Performance and Intermediation Changes Surrounding Sustained Increases in Disclosure. *Contemporary Accounting Research* 16(3): 485-520.

Ho, J. and R. Tang, 2001. Towards an Optimal Resolution to Information Overload: An Infomediary Approach. Proceedings of the International ACM SIGGROUP Conference on Supporting Group Work 2001. September 30 - October 3, 2001, Boulder, Colorado, USA.

Ke, B. and K. Petroni, 2004. How Informed Are Actively Trading Institutional Investors? Evidence from Their Trading Behavior before a break in a String of Consecutive Earnings Increases. *Journal of Accounting Research* 42(5): 895-927.

Ke, B. and S. Ramalingegowda, 2005. Do Institutional Investors Exploit the Post-Earnings Announcement Drift? *Journal of Accounting and Economics* 39: 25-53.

Kothari, S. P., 2001. Capital Markets Research in Accounting. J*ournal of Accounting & Economics* 31: 105-231.

Mashruwala, C., S. Rajgopal, and T. Shevlin, 2006. Why is the Accrual Anomaly not Arbitraged Away? The role of Idiosyncratic Risk and Transaction Costs. *Journal of Accounting & Economics* 42: 3-33.

Merton, R.C., 1987. A Simple Model of Capital Market Equilibrium with Incomplete Information, *Journal of Finance* 42, 483-510.

Petersen, M.A., 2007. Estimating Standard Errors in Finance Panel Data Sets: Comparing Approaches. Working Paper, Northwestern University and NBER.

Pontiff, J., 1996. Costly Arbitrage: Evidence from Closed-End Funds. *The Quarterly Journal of Economics* 111(4): 1135-1151.

Porter, M., 1992. *Capital Choices: Changing the Way America Invests in Industry*. Boston, MA: Council on Competitiveness/Harvard Business School.

Potter, G., 1992. Accounting Earnings Announcements, Institutional Investor Concentration, and Common Stock Returns. *Journal of Accounting Research* 30(Spring): 146-155.

Rock, S., S. Sedo, and M. Willenborg, 2001. Analyst Following and Count-Data Economics. *Journal of Accounting and Economics* 30: 351-373.

Rogers, W.H, 1993. Regression Standard errors in Clustered Samples. *Stata Technical Bulletin* 13 (May): 19–23.

Romer, P. 1990, Endogenous Technological Change. *Journal of Political Economy* 98(5): 71-102.

Serwer, A., 1997. The Scariest Tech Stock Ever! The Ups and Downs of Ascend. *Fortune* 136(9):223-224.

Shankar, S.G. and J.M. Miller, 2006. Market Reaction to Changes in the S&P SmallCap 600 Index. *The Financial Review* 41: 339-360.

Shapiro, C., and H.R. Varian, 1999. Information Rules: A Strategic Guide to the Network Economy. Boston: Harvard Business School Press.

Sias, R., 1996. Volatility and the Institutional Investor. *Financial Analysts Journal* 52(2): 13-20.

Sundarajan, A., 2004. Nonlinear pricing of Information Goods. *Management Science* 50(12): 1660–1673

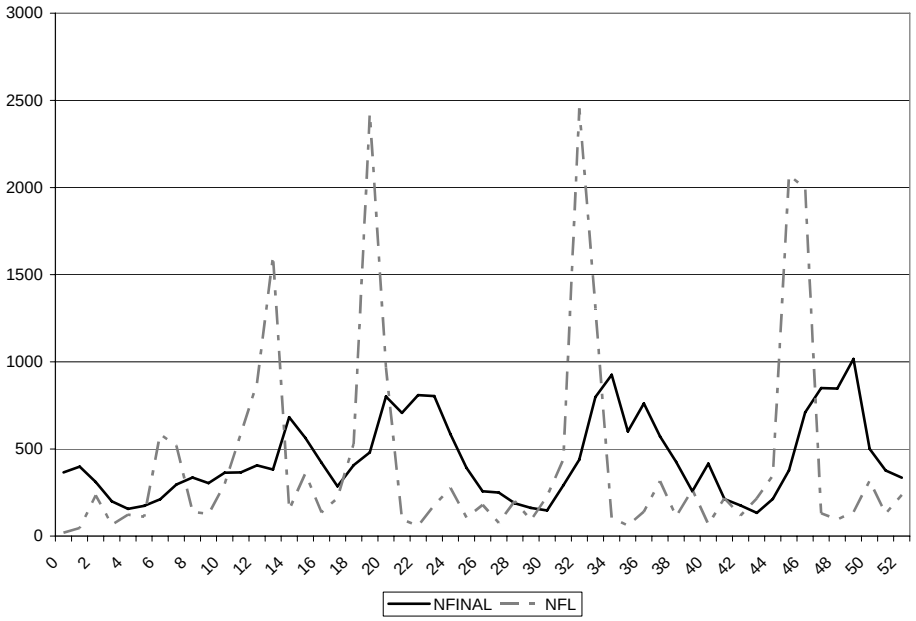
Tees, W., 1992. The Association between Stock Market Responses to Earnings Announcements and Regulation of Electric Utilities. *Journal of Accounting Research* 30(2): 274-285.

Tees, W. and C. Wasley, 1996. Estimating Earnings Response Coefficients: Pooled versus Firm Specific Models. *Journal of Accounting & Economics* 21(3): 279-295.

Williams, R.L., 2000. A Note on Robust Variance Estimation for Cluster-Correlated Data. *Biometrics* 56 (June): 645–646.

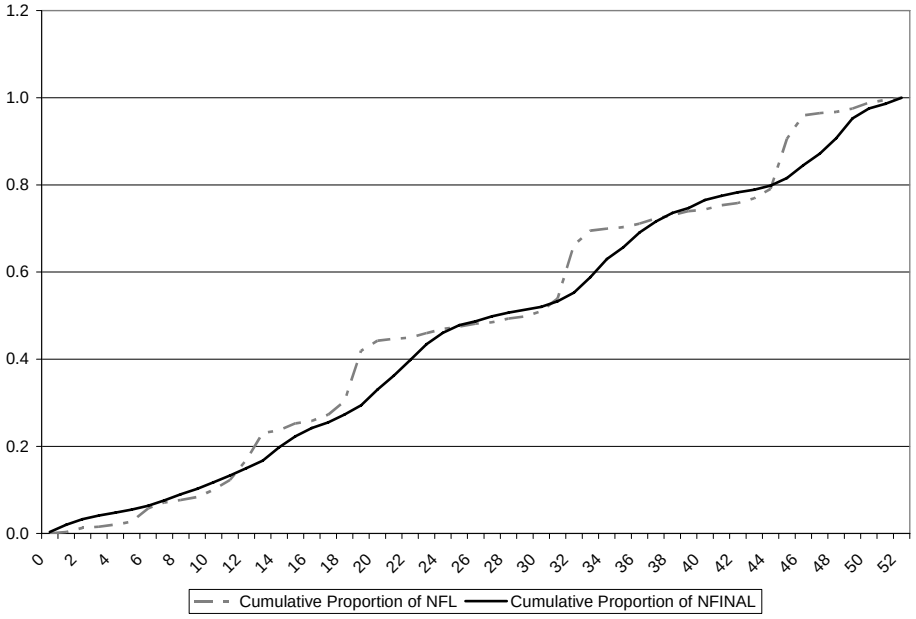
Wurgler, J. and E. Zhuravskaya, 2002. Does Arbitrage Flatten Demand Curves for Stocks? *Journal of Business* 75(4): 583-608.

Figure 1A
Weekly Average Number of SEC Filings and S&P's Finalized Updates, 1991-2004



NFINAL is the weekly average number of final updates completed by S&P over 1991-2004; NFL is the weekly average number of SEC 10-K/10-Q filings over 1991-2004. Due to the measurement error in NFINAL at the beginning of our sample period, we exclude NFINAL and NFL in the first three month (January-March 1991)

Figure 1B
Weekly Average of Cumulative Proportion of 10-K/10-Q Filings and S&P's Finalized Updates, 1991-2004



Solid line is the sum of weekly cumulative number of SEC periodic filings divided by aggregate number of filings over 1991-2004; dashed line is the sum of weekly cumulative number of final updates completed by S&P divided by aggregate number of final updates over 1991-2004. We exclude data in the first three month (January-March 1991) to mitigate the measurement error at the beginning of our sample period.

Figure 2A
Mean difference in FL_CLAG

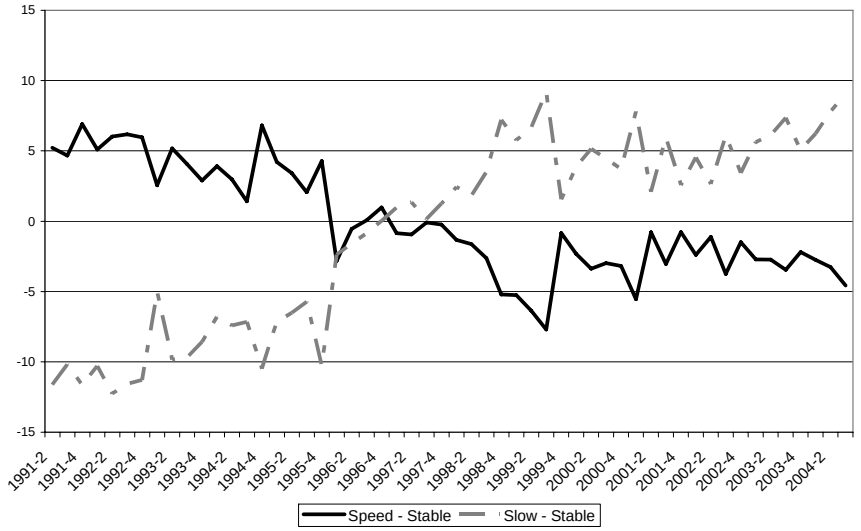


Figure 2B
Mean Difference in INSTOWN

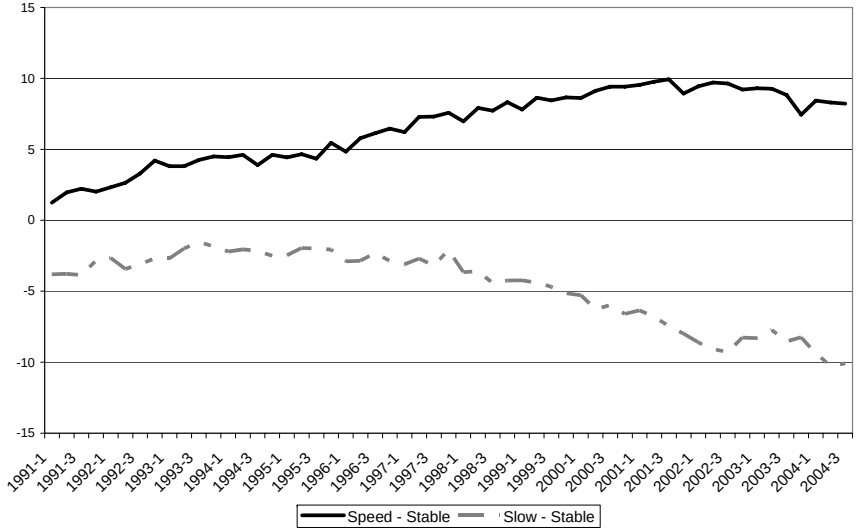


Figure 2C
Mean Difference in S&P500

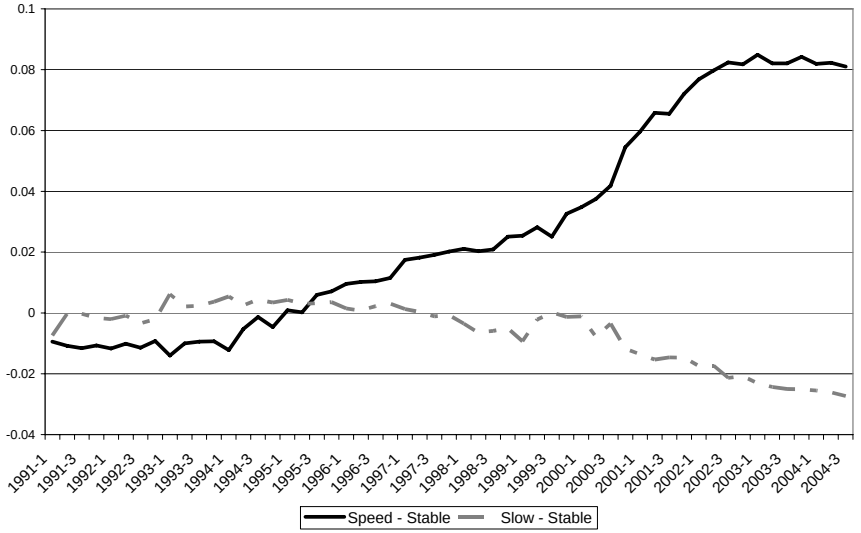


Figure 2D
Mean Difference in CAP1B

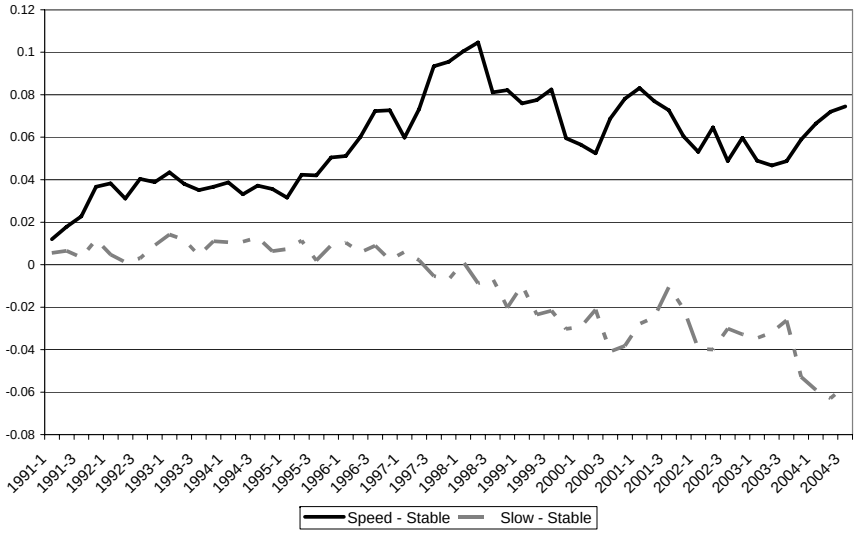


Figure 2E
Mean Differences in S&P 1000

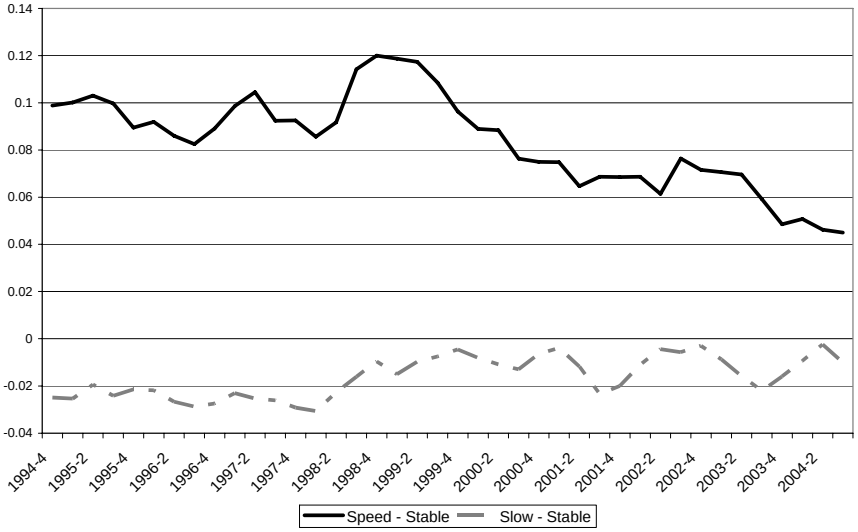


Figure 2F
Mean Difference in VOLUME

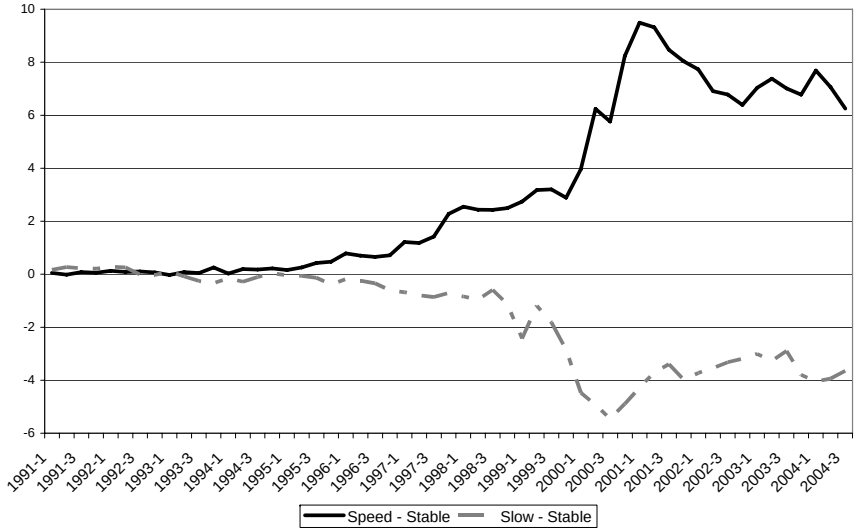
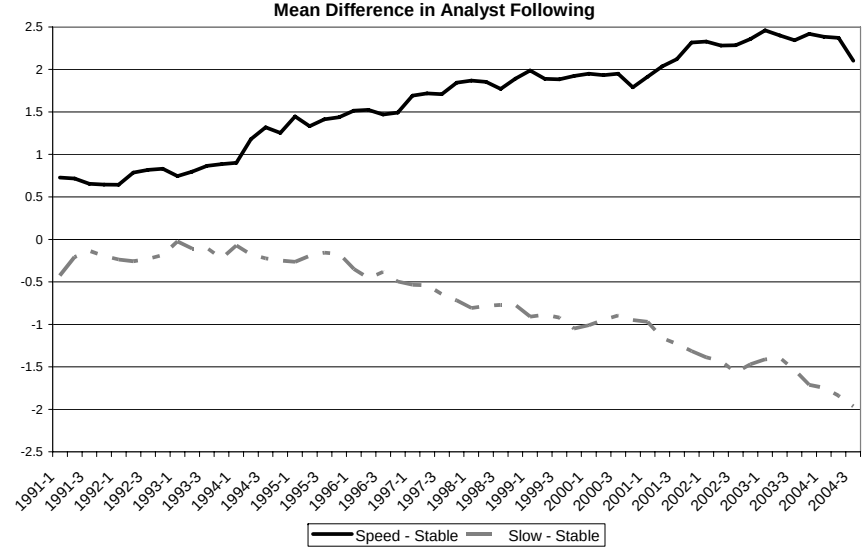


Figure 2G
Mean Difference in CREDITRATE

Figure 2H
Mean Difference in Analyst Following



For each firm, we run a regression as follows:

$$BLOM\_FL\_CLAG_i = \alpha + \beta TREND_i + \varepsilon_i$$

where $BLOM\_FL\_CLAG_i$ is the rank of FL_CLAG. To eliminate the effect of EDGAR on Compustat collection speed, we compute Blom normal scores (Blom 1958) from the ranks. $TREND_i$ is a time trend variable for calendar quarter over 1991Q2-2004Q4 (ranging from 1 to 55). We create two dummy variables, denoted SPEED and SLOW respectively. SPEED (SLOW) equals 1 if β is significantly negative (positive) at the two-tailed level of 5 percent, and 0 otherwise. Finally, STABLE equals 1 if β is insignificant at the two-tailed level of 5 percent, and 0 otherwise. In the above figures, we subtract values of STABLE group from those of SPEED or SLOW group for each variable.

Table 1
Sample Selection and Descriptive Statistics of Computat Data Collection Lags

This table shows our sample selection procedures. We extract Compustat production date from Compustat Preliminary History database, and SEC filing date from SEC filing date database. FL_CLAG is measured as the number of weekdays from a company's 10-K/10-Q filing date to Compustat production date when a company's finalized financial data first appears in Compustat (FINALQPRD).

Panel A. Sample for analysis of periodic SEC filings collection speed	
Firm-quarters in both Compustat Preliminary History, Compustat AFR database, and CRSP over 1991-2004	401,707
Less:	
Firm-quarters with missing SEC filing date or finalqprd	(87,294)
Filing date preceding fiscal quarter end	(49)
Finalqprd preceding filing date	(1,589)
Final Sample	312,775

Panel B: Measures of Speed of Periodic SEC Filing Collection

	N	Mean	Std	Min	Q1	Median	Q3	Max
All	309,654	18.635	15.745	0	8	15	23	131
Pre-EDGAR (1991-1995)	97381	24.736	16.257	0	16	22	29	131
Post-EDGAR (1997-2004)	186552	14.231	12.840	0	7	11	18	131

Table 2
Descriptive Statistics

Panel A: Firm-level demand factors

	N	Mean	Std	Median	Min	Max	Inter-quartile
$INSTOWN_{i,q-1}$	301199	26.249	26.650	17.965	0	100	44.744
$TRAOWN_{i,q-1}$	301199	7.670	11.250	2.301	0	100	11.347
$QIXOWN_{i,q-1}$	301199	12.246	13.860	7.213	0	100	19.882
$DEDOWN_{i,q-1}$	301199	5.699	8.845	0.784	0	100	8.996
$SPECTRUM_{i,q-1}$	301199	0.807	0.395	1	0	1	0
$ARBRISK_{i,q}$	301979	3.955	2.791	3.209	0	19.401	3.054
$PRICE_{i,q}$	310077	16.127	16.718	11	0.0001	102	19.063
$VOLUME_{i,q}$	302312	6.564	24.641	0.345	0	235.893	2.307
$S\&P500_{i,q}$	312775	0.073	0.260	0	0	1	0
$CAP1B_{i,q}$	308097	0.094	0.292	0	0	1	0
$S\&P1000_{i,q}$	308097	0.083	0.276	0	0	1	0
$CREDITRATE_{i,q}$	312745	0.193	0.394	0	0	1	0
$FIINSERV_{i}$	312775	0.190	0.392	0	0	1	0
$UTILITY_{i}$	312775	0.029	0.168	0	0	1	0
$SPECIAL_{i,q}$	312775	0.217	0.412	0	0	1	0
$EXTRAORD_{i,q}$	312775	0.044	0.206	0	0	1	0
$CONCUR_{i,q}$	312775	0.206	0.405	0	0	1	0

Panel B: Macro-level supply factors

	N	Mean	Std	Median	Min	Max	Inter-quartile
$NPRELIM_{p}$	703	0.429	0.384	0.268	0.010.2	2.050	0.538
$NFINAL_{p}$	703	0.445	0.359	0.336	22	2.372	0.425
NEA_{p}	703	0.429	0.385	0.268	0.010	1.680	0.546
NFL_{p}	703	0.444	0.766	0.186	0.010	4.416	0.266
$EA\_BACKLOG_{p-1}$ †	703	0.233	0.216	0.152	0.013	1.263	0.233
$FL\_BACKLOG_{p-1}$ †	703	1.693	1.291	1.216	0.117	7.021	1.717
$QTREND_{p}$	703	0.078	0.269	0	0	1	0
$YEAREND_{p}$	703	0.020	0.140	0	0	1	0

† Given the backlog variables are likely to contain measurement error at the beginning of the sample period, we exclude the first three months' data in our analyses (January-March 1991), but consider all data in the measurement of the backlog variables (beginning January 1991).

Variable Definitions:

$INSTOWN_{i,q-1}$	= percentage of institutional holdings as of the beginning of calendar quarter (winsorized to 100) if SPECTRUM equals 1**,** and 0 otherwise,
$TRAOWN_{i,q-1}$	= percentage of stock ownership held transient institutions at the beginning of calendar quarter (winsorized to 100),
$QIXOWN_{i,q-1}$	= percentage of stock ownership held quasi-indexing institutions at the beginning of calendar quarter (winsorized to 100),
$DEDOWN_{i,q-1}$	= percentage of stock ownership held dedicated institutions at the beginning of calendar quarter (winsorized to 100),
$SPECTRUM_{i,q-1}$	= 1 if institutional ownership data is available from Thomson Financial Spectrum database, and 0 otherwise,
$S\&P\ 500_{i,q}$	= 1 if S&P 500 firms in the current fiscal quarter, and 0 otherwise,
$CAP1B_{i,q}$	= 1 if non-S&P 500 firms with market capitalization greater than 1 billion dollars in the current fiscal quarter, and 0 otherwise,
$S\&P\ 1000_{i,q}$	= 1 if S&P MidCap 400 or SmallCap 600 firms with market capitalization less than 1 billion dollars in the current fiscal quarter, and 0 otherwise,
$ARBRISK_{i,q}$	= standard deviation of residuals from a regression of firm-specific daily returns on the returns of the CRSP equally-weighted market index over the current fiscal quarter (a minimum of 5 observations is required). We multiple the standard deviation of residuals by 100 for presentation purpose and winsorize it at its 99.5 percentile value,
$PRICE_{i,q}$	= close price as of fiscal quarter end, winsorized at its 99.5 percentile value,
$VOLUME_{i,q}$	= dollar value of trading volume in millions averaged over the current fiscal quarter, winsorized at its 99.5 percentile value,
$SPECIAL_{i,q}$	= 1 if a special item is reported in the periodic SEC 10-K/10-Q filings of fiscal quarter q, and 0 otherwise,
$EXTRAORD_{i,q}$	= 1 if an extraordinary item is reported in the periodic SEC 10-K/10-Q filings of fiscal quarter q, and 0 otherwise,
$CREDITRATE_{i,q}$	= 1 if S&P credit rating is available in the current fiscal quarter, and 0 otherwise,
$FINSERV_{i}$	= 1 if financing service firms (SIC 6000-6999), and 0 otherwise,
$UTILITY_{i}$	= 1 if utility firms (SIC 4900-4999), and 0 otherwise,
$CONCUR_{i,q}$	= 1 if earnings announcement date concurs with SEC filing date for the current fiscal quarter, and 0 otherwise.
$NPRELIM_{p}$	= number of earnings press releases (in ‘000) from which *preliminary* data was collected during the production cycle ended at date p,
$NFINAL_{p}$	= number of periodic SEC filings (in ‘000) from which *final* data was collected during the production cycle ended at date p,
$EA\_BACKLOG_{p-1}$	= number of uncollected earnings announcements (in ‘000s) prior to the current production date p,
$FL\_BACKLOG_{p-1}$	= number of uncollected 10-K/10-Q filings (in ‘000s) prior to the current production date p,

NEA_p	= number of earnings announcements (in ‘000s) made during the production cycle ending at date p by all companies followed by Compustat,
NFL_p	= number of 10-K/10-Q filings (in ‘000s) during the production cycle ending at date p by all companies followed by Compustat,
$YEAREND_p$	= 1 if the production date p is the last business day in a year, and 0 otherwise, and
$QTREND_p$	= 1 if the production date p is the last business day in a quarter, and 0 otherwise.

Table 3
Regression of Collection lags on Institutional Ownership and other Demand and Supply Factors[†]

	Predicted Sign	Regression 1			Regression 2		
		Coeff.	*p value*	*Marginal Effect*	*Coeff.*	*p value*	*Marginal Effect*
$INSTOWN_{i,q-1}$	-	-0.0012	0.0000	-5.19%			
$TRAOWN_{i,q-1}$	?				-0.0002	0.2880	-0.25%
$QIXOWN_{i,q-1}$	-				-0.0026	0.0000	-5.22%
$DEDOWN_{i,q-1}$	?				-0.0002	0.2630	-0.22%
$SPECTRUM_{i,q-1}$	?	0.0047	0.3900	0.47%	0.0081	0.1430	0.81%
$ARBRISK_{i,q}$	+	0.0028	0.0000	0.86%	0.0025	0.0000	0.78%
$PRICE_{i,q}$	-	-0.0007	0.0000	-1.33%	-0.0007	0.0000	-1.28%
$VOLUME_{i,q}$	-	-0.0007	0.0000	-0.17%	-0.0008	0.0000	-0.18%
$S\&P\ 500_{i,q}$	-	-0.4014	0.0000	-40.14%	-0.3833	0.0000	-38.33%
$CAP1B_{i,q}$	-	-0.1263	0.0000	-12.63%	-0.1232	0.0000	-12.32%
$S\&P\ 1000_{i,q}$	-	-0.2180	0.0000	-21.80%	-0.2092	0.0000	-20.92%
$CREDITRATE_{i,q}$	-	-0.0440	0.0000	-4.40%	-0.0418	0.0000	-4.18%
$FINSERV_{i}$	+	0.0888	0.0000	8.88%	0.0913	0.0000	9.13%
$UTILITY_{i}$	+	0.2311	0.0000	23.11%	0.2334	0.0000	23.34%
$SPECIAL_{i,q}$	+	0.0181	0.0000	1.81%	0.0180	0.0000	1.80%
$EXTRAORD_{i,q}$	+	0.0691	0.0000	6.91%	0.0680	0.0000	6.80%
$CONCUR_{i,q}$	-	-0.0251	0.0000	-2.51%	-0.0256	0.0000	-2.56%
$RDQEBACKLOG_{p-1}$	+	0.0394	0.0230	1.51%	0.0368	0.0340	1.40%
$RDQEBACKLOG\_SQRD_{p-1}$	-	0.0120	0.2520	0.37%	0.0136	0.1950	0.42%
$FILEBACKLOG_{p-1}$	+	0.1840	0.0000	27.23%	0.1832	0.0000	27.11%
$FILEBACKLOG\_SQRD_{p-1}$	-	-0.0105	0.0000	-5.85%	-0.0103	0.0000	-5.76%
$NRDQE_{p}$	-/?	0.3520	0.0000	4.36%	0.3646	0.0000	4.52%
$NRDQE\_SQRD_{p}$	+/?	-1.0814	0.0000	-2.95%	-1.1142	0.0000	-3.04%
$NFILE_{p}$	-/?	0.1547	0.0000	17.47%	0.1539	0.0000	17.38%
$NFILE\_SQRD_{p}$	+/?	-0.0407	0.0000	-5.99%	-0.0404	0.0000	-5.95%
Y1992	?	-0.1737	0.0000	-17.37%	-0.1724	0.0000	-17.24%
Y1993	?	-0.1188	0.0000	-11.88%	-0.1182	0.0000	-11.82%
Y1994	?	-0.0801	0.0000	-8.01%	-0.0796	0.0000	-7.96%
Y1995	?	0.0021	0.7630	0.21%	0.0019	0.7810	0.19%
Y1996	?	-0.1131	0.0000	-11.31%	-0.1166	0.0000	-11.66%
Y1997	?	-0.5909	0.0000	-59.09%	-0.5954	0.0000	-59.54%
Y1998	?	-0.3156	0.0000	-31.56%	-0.3109	0.0000	-31.09%
Y1999	?	-0.2984	0.0000	-29.84%	-0.2963	0.0000	-29.63%
Y2000	?	-0.6253	0.0000	-62.53%	-0.6223	0.0000	-62.23%
Y2001	?	-0.6977	0.0000	-69.77%	-0.7027	0.0000	-70.27%

(Continued on next page)

[†] For brevity, we suppress the slope estimates of quarter-/year-end dummies and weekday dummies.

Table 3 (Cont'd)

	Predicted Sign	Regression 1			Regression 2		
		Coeff.	*p value*	*Marginal Effect*	*Coeff.*	*p value*	*Marginal Effect*
Y2002	?	-0.9223	0.0000	-92.23%	-0.9289	0.0000	-92.89%
Y2003	?	-0.8581	0.0000	-85.81%	-0.8572	0.0000	-85.72%
Y2004	?	-0.7306	0.0000	-73.06%	-0.7236	0.0000	-72.36%
INTERCEPT	?	2.7940	0.0000		2.7953	0.0000	
Obs			287,515			287,515	
Pseudo R-Squared			0.2268			0.2271	

Table 4
Regression Analysis of the Effects on FL_CLAG from Firms Switching among Major Indices

From \ To (p value) [n]	*NOINDEX*	*S&P 1000*	*CAP1B*	*S&P 500*
NOINDEX		-0.1195 (0.108) [21,518]	0.0255 (0.224) [13,848]	0.0082 (0.802) [25]
S&P 1000	-0.0272 (0.264) [2,622]	-0.0876 (0.2236) [28,315]	-0.1544 (0.000) [10,994]	-0.4443 (0.000) [32]
CAP1B	-0.0097 (0.363) [10,571]	-0.2104 (0.005) [5,934]	-0.0931 (0.000) [36,976]	-0.0583 (0.063) [6,129]
S&P 500	-0.1214 (0.000) [688]	-0.3000 (0.001) [528]	-0.1589 (0.036) [458]	-0.3961 (0.000) [27,478]

This table presents the incremental effect on FL_CLAG when a firm switches from one "index" to another. We estimate these coefficients by adding dummy variables for the incidence of "index" switching in the FL_CLAG regression in Table 3. The results of other variables are omitted for brevity. P-values are included in parentheses, and the number of observations in each switching category in brackets. The coefficients on the diagonal show the collection effects of firms staying in S&P 500, CAP1B and S&P 1000 throughout the entire sample period. The coefficients above (below) the diagonal show incremental effects (relative to the diagonal coefficients) on FL_CLAG when a firm enters from an index with lower (higher) collection priority to an index with higher (lower) collection priority.

Table 5
Regression of institutional ownership (by trading type) on time trend and control variables (two-tailed p value in parentheses) [†, ‡]

		TRAOWN$_{iq}$	*QIXOWN$_{iq}$*	*DEDOWN$_{iq}$*
TREND$_{iq}$	+/+/-	0.1997	0.1014	-0.1265
		(0.000)***	(0.000)***	(0.000)***
TREND_SPEED$_{iq}$	+/+/?	0.0311	0.0295	0.0035
		(0.000)***	(0.002)***	(0.550)
TREND_SLOW$_{iq}$	-/-/?	-0.0270	-0.0136	0.0021
		(0.001)***	(0.251)	(0.741)
MRET$_{iq}$	?	0.1270	-1.9165	-0.1638
		(0.778)	(0.000)***	(0.547)
TVOL$_{iq}$	?	18.5210	9.0062	2.7968
		(0.000)***	(0.000)***	(0.000)***
Ln(MVE) $_{iq}$	?	2.4597	2.9678	1.2920
		(0.000)***	(0.000)***	(0.000)***
BETA$_{iq}$	?	0.2424	0.3364	-0.0866
		(0.037)**	(0.042)**	(0.369)
IRISK$_{iq}$	?	-0.8766	-3.5836	-2.0974
		(0.006)***	(0.000)***	(0.000)***
LEV$_{iq}$	?	1.6920	2.6890	1.1066
		(0.000)***	(0.000)***	(0.000)***
DP$_{iq}$	?	-83.4621	-78.0620	-68.7032
		(0.000)***	(0.000)***	(0.000)***
EP$_{iq}$	?	0.8635	-0.5621	-1.0477
		(0.044)**	(0.108)	(0.000)***
BP$_{iq}$	?	1.0852	1.8108	1.1366
		(0.000)***	(0.000)***	(0.000)***
SGR$_{iq}$	?	-0.0050	-0.0083	-0.0039
		(0.000)***	(0.000)***	(0.000)***
RATE$_{iq}$	?	-0.1149	0.8411	0.0493
		(0.002)***	(0.000)***	(0.198)
S&P500$_{iq}$	?	-3.6784	1.5278	-1.5259
		(0.000)***	(0.026)**	(0.000)***
INTERCEPT	?	-14.4901	-21.4325	-4.5253
		(0.000)***	(0.000)***	(0.000)***
Obs		273,463	273,463	273,463
Adjusted R-Squared		0.3627	0.3813	0.1358

[†] The p values are based on standard errors adjusted for correlations among different firms in the same year and different years in the same firm, i.e., two-way clustering standard errors.
[‡] ***, **, * represent statistical significance at 0.01, 0.05 and 0.1 levels respectively.

Variable Definitions:

$TREND_{iq}$ = a time trend variable for calendar quarter q over 1991Q2-2004Q4 (ranging from 1 to 55). We run a regression for each firm with at least eight quarters data:

$$FL\_CLAG_i = \alpha + \beta TREND_i + \varepsilon_i$$

We create two dummy variables, denoted SPEED and SLOW. SPEED (SLOW) equals 1 if β is significantly negative (positive) at the two-tailed level of 5 percent, and 0 otherwise. $TREND\_SPEED_{iq}$ ($TREND\_SLOW_{iq}$) are interactions of TREND and SPEED (SLOW),

$MRET_{iq}$ = market-adjusted buy-and-hold stock return measured over quarter q (a minimum of 30 observations is required),

$TVOL_{iq}$ = average monthly trading volume relative to shares outstanding at the end of fiscal quarter q,

$Ln(MVE)_{iq}$ = log of the market capitalization of common equity at the end of fiscal quarter q,

$BETA_{iq}$ = market-model beta estimated from daily stock returns over fiscal quarter q (a minimum of 30 observations is required),

$IRISK_{iq}$ = log of the standard deviation of market model residuals estimated from daily stock returns over fiscal quarter q (a minimum of 30 observations is required),

LEV_{iq} = ratio of total debt to assets, both measured at the end of fiscal quarter q,

DP_{iq} = ratio of dividends to market value of equity, both measured at the end of fiscal quarter q,

EP_{iq} = ratio of income before extraordinary items to market value of equity, both measured at the end of fiscal quarter q,

BP_{iq} = ratio of book value of equity to market value of equity, both measured at the end of fiscal quarter q,

SGR_{iq} = percentage change in sales of fiscal quarter q relative to quarter q-1,

$RATE_{iq}$ = S&P stock rating (9 = A+,…, 1 = not rated), and

$S\&P\ 500_{iq}$ = 1 if firm i is in the S&P 500 index, and 0 otherwise.

Table 6
Firm and time fixed effects regression of changes in institutional ownership (by trading type) on S&P collection lag and control variables (two-tailed p value in parentheses) [†,‡]

		$\Delta TRAOWN_{iq}$	$\Delta QIXOWN_{iq}$	$\Delta DEDOWN_{iq}$
$FL\_CLAG\_KNOWN_{i,q-1}$	-/-/?	-0.0216	-0.0130	0.0162
		(0.000)***	(0.000)***	(0.000)***
$UNKNOWN_{i,q-1}$	-/-/?	-0.5798	-0.4124	0.3646
		(0.000)***	(0.007)***	(0.001)***
$OWN_{i,q-1}$	?	-0.2479	-0.2341	-0.2208
		(0.000)***	(0.000)***	(0.000)***
$PORTWT_{iq-1}$	?	-1.1438	-2.4326	-0.0525
		(0.000)***	(0.000)***	(0.052)*
$Ln(MVE)_{iq-1}$	?	0.6993	0.9865	0.0488
		(0.000)***	(0.000)***	(0.288)
BTM_{iq-1}	?	0.0006	0.0026	0.0007
		(0.743)	(0.590)	(0.705)
$RETQ24_{iq}$	?	0.0026	-0.0005	0.0008
		(0.012)**	(0.597)	(0.129)
$RETQ1_{iq}$	?	0.0150	-0.0011	0.0013
		(0.000)***	(0.583)	(0.047)**
$RETQ0_{iq}$	?	0.0093	0.0002	-0.0001
		(0.000)***	(0.875)	(0.928)
UE_{iq}	?	0.6657	0.6426	0.0804
		(0.000)***	(0.000)***	(0.458)
Obs		216,027	216,027	216,027
Adjusted R-Squared		0.1314	0.1189	0.1135

[†] The p values are based on standard errors adjusted for correlations among different firms in the same year and different years in the same firm, i.e., two-way clustered standard errors.
[‡] ***, **, * represent statistical significance at 0.01, 0.05 and 0.1 levels respectively.

Variable Definitions:

In this regression we follow the model by Ke and Petroni (2004). The variables are defined as follows:

$INST_{iq}$ represents $TRAOWN_{iq}$, $QIXOWN_{iq}$ and $DEDOWN_{iq}$ measured as the percentage of ownership at the end of a calendar quarter q by transient, quasi-indexing and dedicated institutions, respectively, and Δ denotes changes in $TRAOWN_{iq}$, $QIXOWN_{iq}$ and $DEDOWN_{iq}$ over a calendar quarter. $INST_{i,q-1}$ is the ownership at the beginning of quarter q for the institution specified in the dependent variable, and $PORTWT_{iq-1}$ is the percentage portfolio weight (weighted by the total market cap of the institution's stock portfolio at the beginning of quarter q) in the portfolio of the institution specified in the dependent variable.

$UNKNOWN_{iq-1}$ equals to 1 if by the end of quarter q the firm have not yet filed their periodic SEC report for the prior quarter or Standard & Poor's has not yet collected the information from the periodic report, and 0 otherwise. $FL\_CLAG\_KNOWN_{iq-1}$ equals FL_CLAG_{iq-1} (zero) when UNKNOWN equals zero (one).

$Ln(MVE)_{iq-1}$ is the natural logarithm of market value of common equity at the beginning of fiscal quarter q. BTM_{iq-1} is the ratio of book value of equity to market value of equity at the beginning of fiscal quarter q. $RETQ24_{iq}$ ($RETQ1_{iq}$) is the buy-and-hold raw return in percentage for two through four calendar quarter (one calendar quarter) prior to the beginning of calendar quarter q. $RETQ0_{iq}$ is the buy-and-hold raw return in percentage from 30 to 3 trading days before the earnings announcement date of quarter q. UE_{iq} is change in earnings before extraordinary items from fiscal quarter q-4 to quarter q.