The Interdependence between Institutional Ownership and Information Dissemination by 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: September 15, 2008

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, Steve Huddart, Tom Lys, Karen Nelson, Santhosh Ramalingegowda (2008 FARS Conference discussant), Frank Selto, Jake Thomas, Jim Vincent, Susan Watts, Frank Zhang, two anonymous reviewers, and the workshop participants at the Columbia University Burton Workshop, George Mason University Corporate Governance Conference (including the discussant, Larry Weiss), Pennsylvania State University, Purdue University, and the University of Colorado at Boulder. We have benefited immensely from discussions with Scott Eston and Chuck Taylor. We are responsible for all remaining errors. We gratefully acknowledge the contributions of I/B/E/S International Inc. for providing analyst forecast data. These data have been provided as part of a broad academic program to encourage earning expectations research. We also thank Brian Bushee for generously providing data on institutional investor categories and discussing his methodology with us.

The Interdependence between Institutional Ownership and Information Dissemination by 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 Ownership and Information Dissemination by 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)

I. Introduction

This paper examines the interrelationship between institutional investor stock ownership and the dissemination speed of key corporate accounting information by Standard & Poor's Compustat. The key role of data aggregators in capital markets is highlighted in the following quote from PWC's letter to the SEC (June 8, 2006) on the use of XBRL:

> "….based on our discussions with investors and analysts, we understand that investors acquire the large majority of relevant information, including financial data, from sources not controlled by the reporting entity. These information intermediaries employ automated parsing technologies and manual resources to parse public filings and other data sources into standardized and electronically reusable data" (http://www.sec.gov/news/press/4-515/4515-8.pdf).

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. We address this this gap by studying the information dissemination practices of Standard & Poor's (S&P, hereafter), an important information intermediary that acts as a key supplier in the market for corporate accounting information.

The sample for our study consists of 319,524 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 while Standard & Poor's firm-specific data collection lag is decreasing in the firm's overall level of institutional ownership, it is also influenced by the nature of institutional investor clientele (Bushee 1998; 2001).[1] Specifically, we predict and find that the relationship between institutional ownership and dissemination lag is driven by the informational demands of quasi-indexers. These results are consistent with Bushee and Noe (2000), who argue that quasi-indexers depend on corporate financial disclosures as a low-cost performance-monitoring mechanism, while neither transient nor dedicated institutions rely on quick access to standardized corporate accounting information to meet their investment objectives.

[1] Bushee (1998, 2001) identifies three types of institutional investors by trading behavior: transient investors are characterized by high diversification, high turnover, and short horizons; dedicated investors by "relation investing" in a few selected firms, with low turnover and long horizons; and quasi-indexers by high diversification, but low turnover.

We also examine several investment characteristics that influence the demand for speedy dissemination of corporate information. Given that high arbitrage risk and transaction costs deter sophisticated investors from exploiting accounting-based mispricing (see Collins et al. 2003; Ali et al. 2003; Mashruwala et al. 2006), demand for corporate accounting information is likely to be lower when arbitrage costs are high. Consistent with this intuition, we predict and find less speedy 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 middle or small cap indices. Results are consistent with hypothesized collection priorities, documenting a 41 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. 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 S&P's 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. Based on extant research (e.g., Bushee and Noe, 2000), we predict that transient investors are more likely to be drawn to firms with faster dissemination of corporate accounting data, since a richer information environment facilitates the implementation of their short-run trading strategies. Since financial disclosures offer a low-cost monitoring mechanism for quasi-indexers, we predict that quasi-indexers will also be drawn to firms with faster information dissemination speeds. Dedicated investors, however, have longer horizons and invest in a few selected firms, and are likely to be indifferent to the liquidity benefits of a richer information environment (Bushee and Noe, 2000). However, there is evidence that dedicated investors who have longer horizons gravitate to investments characterized by higher information asymmetry (Wang and Zhang, 2006), consistent with incentives to retain their informational advantage. Given the ambiguity in the expected behavior, we make no directional predictions about the effect of dissemination speed on the trading decisions of dedicated investors.

We test the impact of dissemination speed on institutional ownership by analyzing the quarter-to-quarter changes in ownership by transient investors and quasi-indexers as a function

of the speed with which corporate accounting information is disseminated by S&P, after controlling for various determinants of changes in institutional ownership (Ke and Petroni, 2004; Ke and Ramalingegowda, 2005). For both transient investors and quasi-indexers, results are consistent with predictions: these institutional investors increase (decrease) their quarter-to-quarter holdings as information dissemination speed increases (decreases). Although we made no specific predictions for dedicated investors, results indicate that these investors increase their ownership position when dissemination of data from periodic filings is delayed, consistent with arguments in extant research that they prefer coarser information environments.

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, and 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 II provides a brief description of the theory regarding the market for information goods, and discusses our hypotheses. Section III discusses sample selection and descriptive evidence. Section IV examines the determinants of the dissemination speed of corporate accounting information, followed by an analysis of the consequences of dissemination speed in section V. The results of additional sensitivity analyses are discussed in Section VI. Concluding remarks are provided in the final section.

II. Theory and Hypotheses Development

In this section we first describe the unique characteristics of the market for corporate accounting information, with special focus on the commercial market segment served by S&P's Compustat. We conclude this section with a discussion of our hypotheses.

II.1 Market for Information Goods: the Case of Corporate Accounting Information

Shapiro and Varian (1999, p.23, emphasis in original) suggest that "the first and most important point is that the markets for information will not, and *cannot*, look like textbook-perfect competitive markets in which there are many suppliers offering similar products, each lacking the ability to influence prices… [Competitive market structure] simply isn't viable for a good in which the incremental cost of production is zero." Consequently, one would expect that information goods would be traded in a differentiated product market, where "a number of firms [are] producing the same 'kind' of information, but with many different varieties." (p. 25) Overall, the market for information goods faces an oligopolistic competition with a handful of large players providing differentiated products and services.[2]

Consistent with the academic intuition on the market for information goods (Shapiro and Varian, 1999), capital market data aggregators 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), the data services are available only to paying customers. 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 the data provided by

[2] We reviewed the periodic filings of a pure play entity in the market for corporate financial information (i.e., FactSet) to identify key competitors in this market. The 2007 10-K report of FactSet identifies only three corporate financial data aggregators as competitors, i.e., Bloomberg, S&P, and Thomson-Reuters. Consistent with FactSet's description of the competition, our discussions with a large mutual fund indicate that, for purposes of quantitative analysis, they rely on Compustat, Worldscope (Thomson), IBES (Thomson), and Ford Equity Research for corporate financial information (with CRSP acting as the primary source for stock market data). As of now, this market has not been significantly impacted by the instant availability of data through electronic tagging made possible by EDGAR. Consequently, data aggregators such as Edgar Online and 10-K Wizard do not yet appear to be direct competitors of S&P, Bloomberg, and Thomson-Reuters.

these aggregators are used are not likely to depend on instantaneous access, suggesting that customers neither demand nor are willing to pay for immediate data access.

Many sophisticated market participants demand the services of data aggregators such as Compustat.[3] For instance, the main subscribers of Compustat's commercial database include buy-side hedge funds, large and small money managers, analysts, quantitative researchers, and corporations. In addition, Compustat charges subscription fees based on assets under management to encourage broader commercial appeal.[4] Our discussions with S&P and a large money manager suggest that institutional investors who engage in sophisticated statistical analysis of accounting information for their investment strategies (akin to "quasi-indexers") require high-quality standardized data that is not only comprehensive in the cross-section, but also have a long time-series. A comprehensive coverage of public companies gives Compustat an advantage in assisting money managers to develop unbiased strategies. The long time-series is especially attractive to money managers who need to back test using historical data for any new investment strategies they develop.

While instantaneous access is not universally demanded, a vital attribute of each aggregator's perceived service quality is likely to be the speed with which it provides information most in demand by its customer base. However, data aggregators cannot "stock" information in anticipation of customer demand. To enhance dissemination speed, aggregators

[3] The discussion that follows is based on the four steps we took to learn about the demand for Compustat's products and services: (1) review of S&P's marketing brochures and pamphlets; (2) participation in a Webex session offered by S&P to potential subscribers; (3) a phone interview with a Compustat staff member; and (4) various e-mail exchanges with Compustat staff.

[4] This pricing scheme is consistent with the increasing value of information as a function of assets under management. Given that the differential commercial price arrangements used by Compustat do not lead to different timeliness of data delivery among commercial customers, we view commercial customers as one market segment for purposes of this paper. 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 is non-trivial.

build up costly production capacity[5] to cater to both client-specific requests for expedited information on individual companies and the general demand from their institutional equity clients. We, therefore, argue that a key choice variable for each data aggregator is its production capacity deployment strategies as reflected in firm-specific dissemination speed choices.

II.2 Hypotheses Development

In the following two subsections we develop hypotheses about 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.

II.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 that stocks held by institutional owners will receive higher priority in general. In addition, our discussions with Compustat indicate that special requests by clients receive the *highest* data collection priority, and Compustat expedites data collection efforts in such cases.[6] Such special requests are likely to further speed up information dissemination for 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.

[5] For instance, to robustly convert the non-standard formats in which corporate accounting data are provided into a standardized electronic database, S&P's Compustat hires data collectors with accounting expertise and trains these professionals in detailed standardization and coding procedures. The Compustat User's Guide is over 700 pages long, with detailed explanations of the specific information that is included in or excluded from each data item.
[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.

Bushee and Noe (2000) posit that quasi-indexers prefer firms with forthcoming disclosures as they offer "a cost-effective method of monitoring firm performance" (p. 176). Consistent with this argument, they find that quasi-indexers gravitate more towards stocks of firms with higher analyst ratings of their mandatory disclosure practices than to those with stronger investor relationship programs. Given that quasi-indexers appear to depend on corporate accounting disclosures as a low cost performance monitoring mechanism, S&P is likely to expedite its data collection efforts for stocks favored by them.

Dedicated institutions follow a relationship strategy with low portfolio turnover and large stakes in the smaller group of stocks they follow (Porter 1992; Dobrzynski 1993). In addition, they rely less on public disclosures as they have better access to private information about their portfolio firms (Porter 1992; Bushee and Noe 2000). Not only is public disclosure less important for their monitoring activities, dedicated institutions might even consider them "potentially costly if it reveals proprietary information" (Bushee and Noe 2000, p. 176). As dedicated investors rely less on public channels of communication, data aggregators are likely to place a lower priority on stocks held by these institutional investors.

While the trading decisions of transient institutions could be indirectly affected by the dissemination speed (see Section II.2.2 for investor response predictions), Bushee and Noe (2000) argue that these institutions do not necessarily *rely* on the accounting information provided by information intermediaries, which in turn suggests that they are unlikely to be a significant source of demand for firm-specific information dissemination speed. 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]

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 indicate that S&P's data collection priorities are designed to reflect the greater demand for information on more visible stocks, which leads 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]

---

[8] We interviewed the chief operating officer and a member of the quantitative research team of an international investment management firm with client assets close to $150 billion. Our discussions with them are entirely consistent with H2. Their quantitative research portfolio (akin to quasi-indexers) staff employs data-driven statistical analysis that relies heavily on publicly available aggregated data. On the other hand, their relational investment ("dedicated investor") portfolio staff closely monitors companies' financial performance (including reviewing filings, press releases, other market information, meeting with management, etc.), but do not rely on standardized panel data provided by data aggregators.

[9] See Bos and Ruotolo (2000) for the criteria for inclusion of a stock in one of the S&P domestic indices.

[10] Compustat staff indicated that their data collection and processing priority is as follows: client requests, S&P 500, TSE 300, billion dollar market cap, remaining S&P 400/600, remaining Russell 3000, and low priority. While S&P

II.2.2 The Institutional Investor Response to Dissemination Speed

Our final hypothesis 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 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. They suggest that transient investors value "more forthcoming disclosure practices because such practices lessen the price impact of trades, facilitating the realization of short-term trading gains" (p. 173). In other words, a finer information environment preserves the liquidity in the market for a stock, which is "likely the key benefit of disclosure for transient institutions" (p. 187). A 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 to be influenced by the dissemination speed to the extent it impacts market liquidity. Therefore, 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.

The results in Bushee and Noe (2000) suggest that "quasi-indexers are more sensitive to the informativeness of mandatory reports than to investor relations activities, which is consistent with their fragmented ownership positions and reliance on corporate disclosures to monitor firm performance" (Bushee and Noe 2000, p. 187). Given that corporate financial disclosures offer a

---

collection priorities may have evolved over time, H3b captures their essence in more recent times. We do not create a separate category for remaining Russell 3000 stocks or stocks with less than or equal to $250 million, so their collection priority is reflected in the intercept. For sake of parsimony, we also do not create a separate category for TSE 300, which represents just over 1.5 percent of sample firm-quarters. In addition, 61 percent of these firm-quarters are in the S&P 1500 index (including S&P LargeCap 500, MidCap 400 and SmallCap 600) or have market capitalization greater than 1 billion, so most TSE firms would be represented in the categories described in H3b. While the indicated collection priority is relevant during our sample period, it appears that Compustat has since shifted the data collection priority of S&P 400/600 firms above that of firms with market capitalization greater than $1 billion (http://www2.standardandpoors.com/spf/pdf/products/Compustat2006.pdf).

low-cost monitoring mechanism for quasi-indexers, we also predict that quasi-indexers will be drawn to firms with faster information dissemination speed. This provides the rationale for our prediction that quasi-indexers both *demand* faster dissemination of corporate accounting disclosures (H2) as well as *gravitate* towards stocks whose accounting information is disseminated quickly. Unlike transients, quasi-indexers prefer finer information environment for the sake of the information itself rather than necessarily the higher market liquidity created by it.

Using equity analyst ratings as a proxy for public disclosure quality, Bushee and Noe (2000) find that "dedicated institutions… show no sensitivity to disclosure rating levels or changes, suggesting that corporate disclosure practices are not a significant factor affecting these institutions' investment decisions" (Bushee and Noe 2000, p. 173). In addition, due to their aversion to frequent trading, dedicated institutions are unlikely to value the liquidity benefits of a finer information environment (Bushee and Noe 2000, p. 176). They are less dependent on public channels of communication, and might arguably even prefer disclosure environments that enable them to retain their informational advantages. Consistent with this possibility, Wang and Zhang (2006) find that the equity holdings of dedicated institutions "are positively associated with indicators for information asymmetry." The implication in the literature is that these institutions are long-term focused, so they are less interested in short-term price movements, but expect to benefit from the market "ultimately" recognizing the value of their holdings. Given the ambiguity in the expected response, 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:

> H4: Current changes in transient and quasi-indexer ownership are negatively associated with prior corporate accounting information dissemination lags.

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

III.1 Sample Identification

Our sample selection procedure is summarized in Panel A of Table 1. Given that we examine information diffusion by capital market intermediaries, we focus on publicly-traded firms in the CRSP database, and obtain our sample from the intersection of CRSP, the quarterly Compustat Preliminary History (Prelim) database, the quarterly Compustat "As First Reported" (AFR) database, and S&P's SEC filing date database. Based on the above considerations, we begin with a sample of 329,048 firm quarters during the calendar years 1991 through 2004. We exclude observations with missing production dates for the delivery of SEC filing information, or with collection lags greater than the 99th percentile. A small set of observations is dropped due to possible data coding problems. The resulting SEC filing sample is 319,524 firm quarters.

III.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 dissemination lag (*DL*) as the number of weekdays between the SEC periodic report filing date and *FINALQPRD*.

Panel B of Table 1 indicates that the median *DL* is 15 weekdays (equal to Compustat's timeliness standard for AFR data during peak season) with the inter-quartile value ranging from eight to 23 weekdays.[12] We also provide descriptive statistics on *DL* separately for the pre-

[11] See http://www.compustat.com/support/wi/private_shrd/dataguide/finalqprd.html.
[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. Compustat offers its

EDGAR (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 coefficient of variation increased from 0.65 to 0.89 after EDGAR, which suggests a large cross-sectional variation in dissemination speed even in the electronic filing era.

III.3 Compustat's Collection Efforts by Calendar Time

In Figure 1A we plot the average number of SEC filings (*#FL*) and "final updates" (*#FL_UPDATE*) (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). When compared to the arrival of information (or filing of periodic reports), S&P's collection efforts are more evenly distributed in calendar time. In Figure 1B we plot the average cumulative proportion of SEC filings and their collection over the 52 production weeks. 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 databases that provide comprehensive accounting information on a broad cross-section of publicly-traded companies

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, note that *DL* 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.

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

IV.1 Model Development

To test our hypotheses on the determinants of information dissemination speed, we estimate the following panel-data regression:[13]

$$
\begin{aligned}
DL_{iq,tp} &= \beta_0 + \beta_1 INST_{iq-1} + \beta_2 SPECTRUM_{iq-1} + \beta_{3a} ARBRISK_{iq} + \beta_{3b} PRICE_{iq} + \beta_{3c} VOLUME_{iq} \\
&+ \beta_{4a} S\&P500_{iq} + \beta_{4b} CAP1B_{iq} + \beta_{4c} S\&P1000_{iq} + \delta_5 RATING_{iq} + \delta_{6a} FINSERV_i + \delta_{6b} UTILITY_i \\
&+ \delta_{7a} SPECIAL_{iq} + \delta_{7b} EXTRAORD_{iq} + \delta_8 CONCUR_{iq} + \delta_{9a} EA\_BKLG_{t-1} + \delta_{9b} (EA\_BKLG_{t-1})^2 \\
&+ \delta_{10a} FL\_BKLG_{t-1} + \delta_{10b} (FL\_BKLG_{t-1})^2 + \delta_{11a} \#EA_t + \delta_{11b} (\#EA_t)^2 + \delta_{12a} \#FL_t + \delta_{12b} (\#FL_t)^2 \\
&+ \delta_{13} YREND_p + \delta_{14} QTREND_p + \delta_{15} MONDAY_p + \delta_{16} TUESDAY_p + \delta_{17} THURSDAY_p + \delta_{18} FRIDAY_p \\
&+ \sum_{T=1992}^{2004} \beta_{19,T} YT_{iq} + \varepsilon_{iq,tp}.
\end{aligned}
\tag{1}
$$

All variables used in this study are defined in the appendix. We use the overall percentage of institutional ownership (*INST*) in the first version of the model (to test H1) and break it down into equity ownership by transients, quasi-indexers and dedicated investors in the second version (to test H2).[14] 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

[13] 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 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.

[14] We obtain the 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 the institutional ownership of the calendar quarter immediately preceding the fiscal quarter q. Brian Bushee kindly provided the yearly classification of institutional investors into transient, quasi-indexers, and dedicated INVESTORS. We conduct a battery of robustness tests and find that Bushee's classification groups exhibit characteristics consistent with the posited investment styles. We also conduct a set of sensitivity analyses of the classification method on our hypotheses tests and find our inferences remain unchanged. Details of these analyses are available from the authors.

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. 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 and utilities, 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.[15] To examine any crossover effects of earnings

[15] 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

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 the incidence of special or extraordinary items, as firms reporting these items are likely to experience business events that influence data aggregation complexity. To maintain its quality standards, Compustat is likely to take longer to collect data for such firms. We also include controls for instances where earnings announcements and filing dates coincide, and calendar-year dummies to examine the supply side effect of EDGAR on information dissemination.

IV.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 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

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.

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 *#EA_UPDATE* and *#EA* 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 *#FL* (0.781) is more than double that of *#FL_UPDATE* (0.367), indicating a smoother inter-temporal collection effort for filings, and suggesting a higher priority given to the collection of earnings announcements.

IV.3 Regression Results

Table 3 shows the results of estimating model (1) using Poisson regression, using institutional ownership in the aggregate as well as broken down by investment style. Except when examining the institutional category 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 forthcoming). 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 4.58 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, consistent with H2. The faster data collection by Compustat suggests that 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 (41 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.

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 22 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 insights on the supply-side effects as well. Compustat's collection lag increases by seven 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 effect of information complexity on dissemination lag is further corroborated by the significantly positive coefficient on *SPECIAL*, indicating that collection lag increases when firms report special items in their filings. 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 (linear term) 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 (quadratic term). 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. 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, consistent with the EDGAR effect becoming more pronounced beginning 1997.

IV.4 Interactions between investor types and other determinants of dissemination lag

We next consider the possibility of an interaction between investor types and other key determinants of dissemination lag, specifically the variables used to test H3a and H3b. First, we construct a combined measure of overall arbitrage costs, encompassing both arbitrage risk and transaction costs. We sort *ARBRISK* (ascending), *PRICE* (descending) and *VOLUME* (descending) into terciles ranked 0 to 2, and add the rankings to get a firm-specific combined score ranging from 0 to 6. We create three dummy variables indicating low, medium or high arbitrage costs when the combined score is 0-1, 2-4, or 5-6, respectively. We interact *TRA*, *QIX* and *DED* with these three dummies and re-estimate the Table 3 regression. Interaction coefficient results are presented in panel A of Table 4. Consistent with sophisticated investors shunning stocks with high arbitrage costs (Collins et al. 2003; Ali et al. 2003; Mashruwala et al. 2006), we find that the interaction coefficients for quasi-indexers indicate a monotonically declining demand for speedy delivery of Compustat data as arbitrage costs increase. None of the interactions is significant for transients. For dedicated investors, the interaction coefficient is significantly positive for stocks with low arbitrage costs, consistent with a lack of demand for information. However, for stocks with high arbitrage cost, the coefficient is significantly negative, consistent with dedicated investors demanding information on stocks with higher arbitrage costs and potentially higher information asymmetry (Wang and Zhang 2006).

We also interact institutional ownership measures with three dummies corresponding to membership in the S&P 500, billion dollar market cap, and S&P 400/600 indices, and a fourth dummy indicating no membership in any market index,, and present interaction coefficients in

Panel B of Table 4. Given the heightened investor awareness for index stocks (Wurgler and Zhuravskaya 2002; Chen et al. 2004), one would expect greater institutional demand for stocks included in major market indices. Results for quasi-indexers indicate that the collection lag is the shortest for S&P 500 firms, increases for stocks in less prominent indices, and is the longest for stocks that do not belong to any major index. The results for transients do not lend themselves to an intuitive explanation. Similar to our results for arbitrage costs, we find evidence consistent with dedicated institutions demanding information on stocks with potentially lower investor awareness.

IV.5 Effect of index switching on dissemination speed

To the extent changes in investor awareness is asymmetric to additions and deletions from major market indices (Chen et al. 2004), 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. Untabulated results indicate that there is a monotonic increase in the dissemination speed as we move from the lowest (S&P 1000) to the highest (S&P 500) priority group for firms that remained in the same priority group throughout our sample period. In terms of "index" switching, when firms are first added to the lowest priority index (S&P1000), S&P's collection lag decreases by 13 percent with an additional decline of 17 percent when their market capitalization goes above $1 billion ("large firm"), and another 5.7 percent reduction when they move to S&P 500. However, consistent with the investor awareness literature, the dissemination speed increases for large firms only if their stock was first part of a major index. 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 13 percent shorter than other firms without an "index" membership (see Chen et al., 2004). Consistent with Shankar and Miller (2006) S&P does not retain any dissemination priority for stocks after they exit the S&P 1000 index.

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

In section V.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 H4 are discussed in section V.2.

V.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 *DL* 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 *DL* 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. The directional evidence in Figure 2A shows, by construction, two sets of firms with divergent raw 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.

V.2 Institutional Investor Response to Changes in Information Environment

Following Ke and Petroni (2004) (see also Ke and Ramalingegowda, 2005), we estimate the following model to examine the predictions of H4 regarding the institutional investor response to dissemination speed:

$$\Delta INST_{iq} = \beta_0 + \beta_1 DL\_KNOWN_{iq-1} + \beta_2 INFOSTATUS_{iq-1} + \delta_3 OWN_{iq-1} + \delta_4 PORTWT_{iq-1} + \delta_5 Ln(MVE_{iq-1}) + \delta_6 BTM_{iq-1} + \delta_7 RETQ24_{iq} + \delta_8 RETQ1_{iq} + \delta_9 RETQ0_{iq} + \delta_{10} UE_{iq} + \varepsilon_{iq}. \quad (2)$$

The dependent variable in model (2) 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 (*DL*) 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 H4. Unreported firm-specific analysis suggests that, on average, *DL* does not exhibit significant inter-temporal persistence. After subtracting its cross-sectional mean to control for the EDGAR effect, we conduct a firm-specific regression of *DL* on its one-period lagged value and find that the median slope is 0.070 with 50 percent of observations falling between -0.095 and 0.254. Given the lack

of an inter-temporal persistence in *DL*, 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 (*INFOSTATUS*) to indicate these cases. We then define the collection lag variable *DL_KNOWN* as equal to the actual collection lag *DL* (zero) when *INFOSTATUS* equals zero (one). Based on H4 we predict a negative slope coefficient for *DL_KNOWN* in the models with changes in transient institution and quasi-indexer ownership as the dependent variable. With respect to *INFOSTATUS*, while we do not make specific predictions, the slope would have the same sign to the extent *INFOSTATUS* 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 effects for firms and for time periods.

The results of estimating model (2), with standard errors corrected for heteroskedasticity and clustering by both firm and time, are reported in Table 5. Consistent with H4, the coefficient of *DL_KNOWN* is significantly negative in the *ΔTRA* and *ΔQIX* models. For a shift in dissemination speed equal to its inter-quartile range (15 weekdays) the short-run effects account for around 0.24 (0.26) percentage points of the quarterly change in transient (quasi-indexer) institutional ownership. While we do not make specific predictions for dedicated institutions, we 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 *INFOSTATUS* has the same sign as that of the collection lag variable, but is not statistically significant for transient investors.[16]

The slope estimates for the control variables are largely consistent with prior studies. 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 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.[17]

VI. Additional Sensitivity Analyses

VI.1 Alternative measures for institutional ownership

We also re-ran all analyses using alternative measures of institutional ownership. We substituted the natural logarithm of the number of investors of each type that hold a stock at the end of each quarter for *INST*, *TRA*, *QIX*, and *DED* and report the results in Table 6. For the Table

[16] Given that *INFOSTATUS* is more likely to take the value of one during the fourth fiscal quarter than during the interim quarters, we modify model (2) and estimate separate slope coefficients for *INFOSTATUS* for interim and fourth quarters. The estimates and the statistical significance for the two coefficients are very similar.

[17] As a robustness check, we include additional control variables from Bushee and Noe (2000, Table 4) that are not considered in (2) (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 *DL_KNOWN* and lagged *INFOSTATUS* and find that the former is insignificant in any model, while the latter is statistically significant with the correct sign in all models.

3 analysis, while both number of transient and quasi-indexer investors influence Compustat collection speed, we still find that quasi-indexers play the dominant role (the difference between the two estimates is statistically significant). The significantly positive coefficient on *DED* is consistent with S&P shunning stocks held by dedicated investors. For the Table 5 analysis, our results remain unchanged using this alternative measure. Results (untabulated) remain unchanged when we consider only the number of institutions with greater than 5% stock holdings.

VI.2 Excluding banks

Institutions classified as banks in the 13-f database hold, on average, 4.2 percent of the shares outstanding of our sample firms. A large majority (more than 75 percent) of these holdings are concentrated in banks classified as quasi-indexers. Give banks face the strictest fiduciary standards for purposes of investment (Bushee 2001), we re-ran all analyses excluding banks to test the robustness of our results. The results are presented in Table 7. In the Table 3 analysis, we find that the holdings of all three categories of non-bank institutions are negatively associated with dissemination lag. However, the magnitude of the slope coefficient for quasi-indexers is at least three times as large as each of the other two coefficients, and the differences are statistically significant (untabulated): demand for dissemination speed is largely driven by the needs of quasi-indexers. Table 5 results remain unaltered when we exclude banks from the sample.

VI.3 Simultaneous estimation

We test our demand determinants hypotheses (H1 through H3) and our consequences hypothesis (H4) through models (1) and (2) (Tables 3 and 5). In both models, the endogenous variables (dissemination lag, institutional ownership) are not hypothesized to influence each

other contemporaneously. Nonetheless, to alleviate possible simultaneity concerns, we estimate (1), with institutional ownership broken down by type, and the three equations represented by (2) jointly using three-stage least squares (3SLS).[18] Untabulated results are qualitatively similar to those reported earlier, and all major inferences hold. One difference is that the slope coefficients of *TRA* and *DED* in (1) become significantly positive. Additional analysis indicates that this result is driven by the use of log transformed linear regression instead of a Poisson regression.

VI.4 Fama-MacBeth regressions

As an additional sensitivity measure, we ran the regressions separately by calendar year-quarter, and computed each test statistic as the mean coefficient divided by its time-series standard error. Untabulated results in Table 3 and Table 5 are qualitatively unchanged in most instances. In Table 3, the coefficient of *DED* becomes significantly negative and that of *ARBRISK* becomes insignificant. All major inferences of Table 5 are robust to the use of the Fama-Macbeth approach.

In summary, our hypotheses test results and sensitivity analyses 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. Overall, our

[18] We are not aware of any statistical package that can jointly estimate Poisson (1) and OLS (2) as a system. To circumvent this problem, we convert *DL* (the dependent variable in (1)) to its natural logarithmic value.

findings indicate that not only do the data aggregators shape the information environment of firms, their dissemination choices also affect investor trading behavior.

VII. 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 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.

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

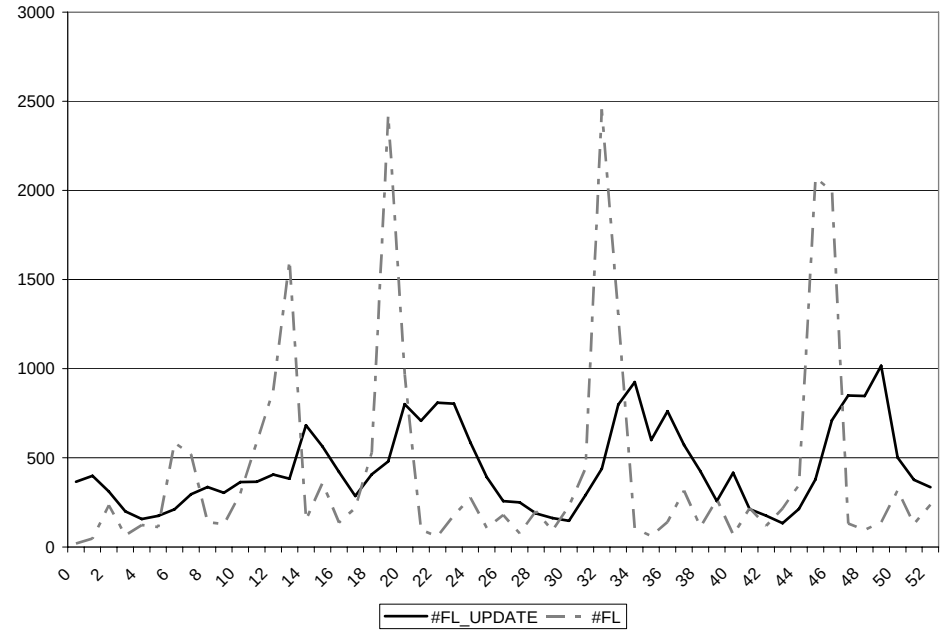
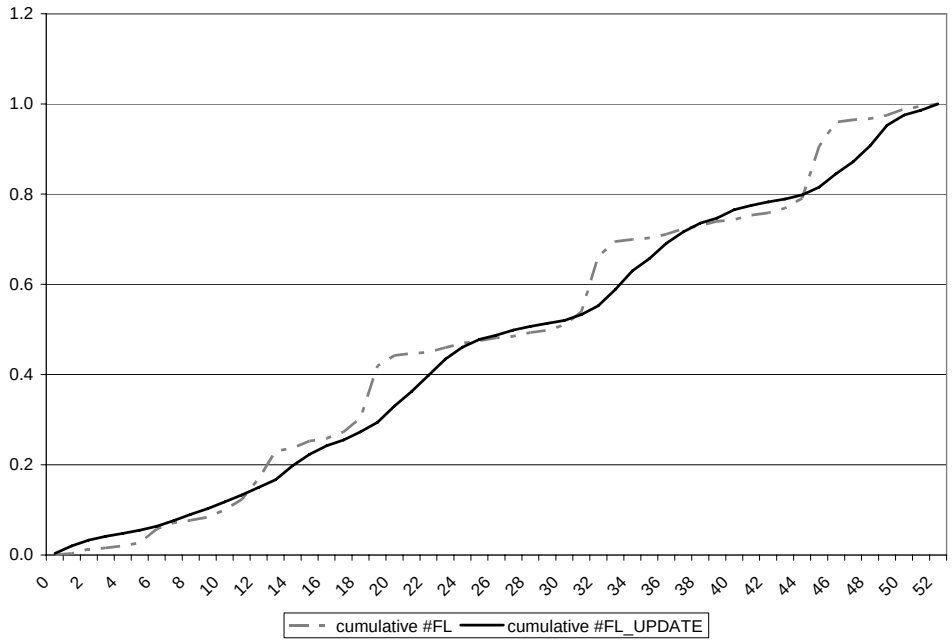


Figure 1B
Weekly Average of Cumulative Proportion of Releases of 10-K/10-Q Filings and S&P's collections, 1991-2004



The above figures compare the intensity of corporate financial information releases versus S&P data collection efforts over 1991-2004. In Figure 1A, #FL is the weekly average number of SEC 10-K/10-Q filings, while #FL_UPDATE is the weekly average number of S&P's collections of SEC filings data, both averages calculated using all weekly data over the sample period. In Figure 1B, we divide the cumulative #FL by the aggregate number of final updates (solid line), and divide cumulative #FL_UPDATE by aggregate number of S&P's final updates (dashed line). Due to the measurement error in #FL at the beginning of the sample period, we exclude data in the first three month (January-March 1991).

Figure 2A
Mean difference in DL

Figure 2B
Mean Difference in INST

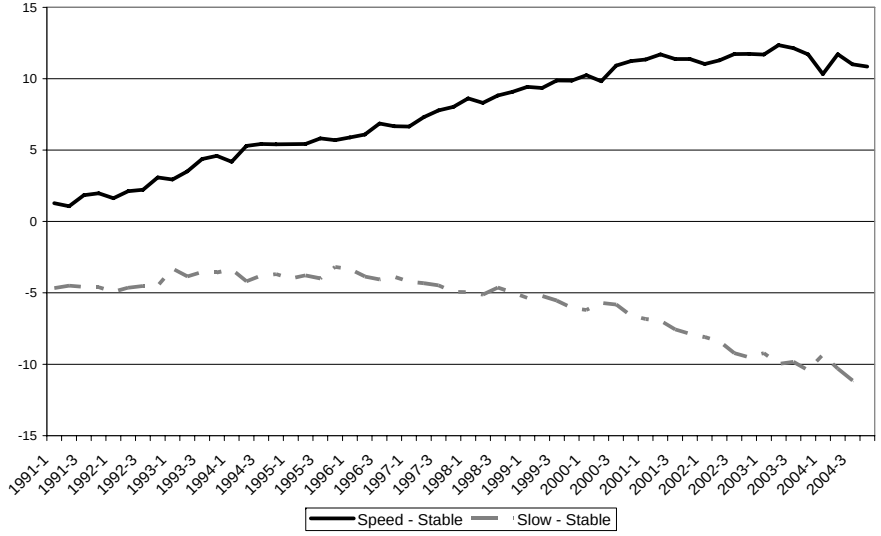


Figure 2C
Mean Difference in S&P500

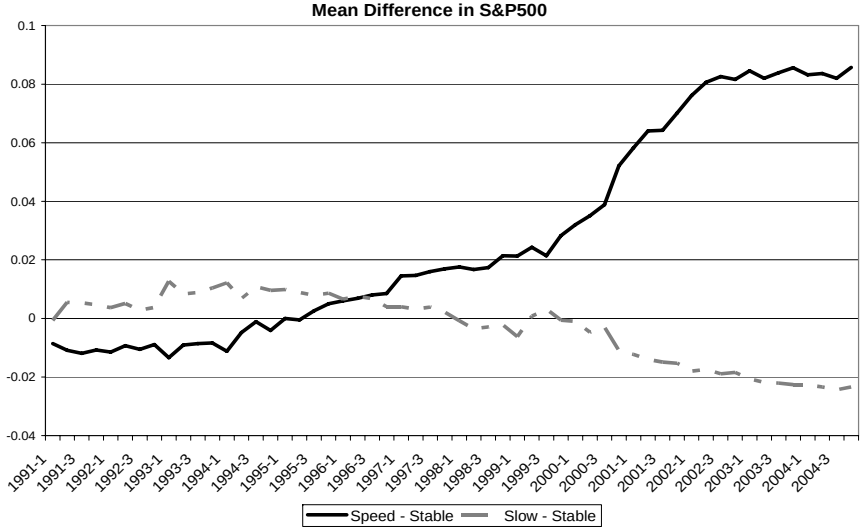


Figure 2D
Mean Difference in CAP1B

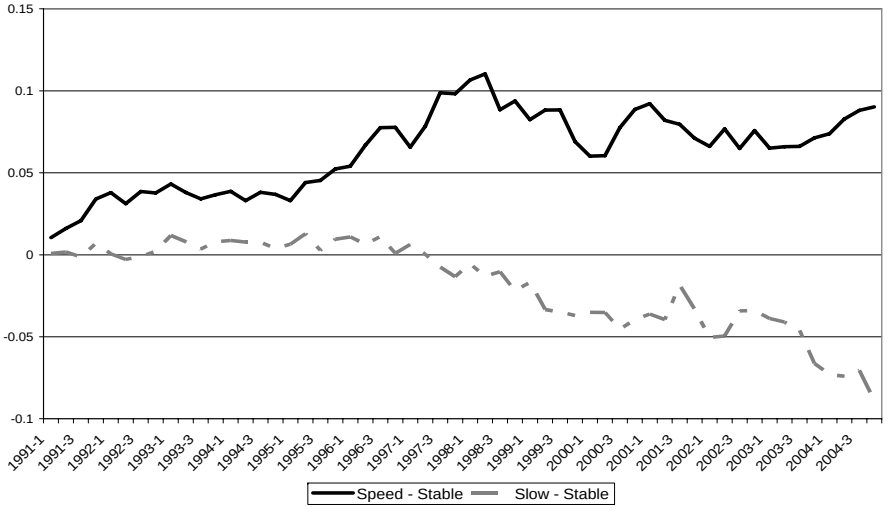


Figure 2E
Mean Differences in S&P 1000

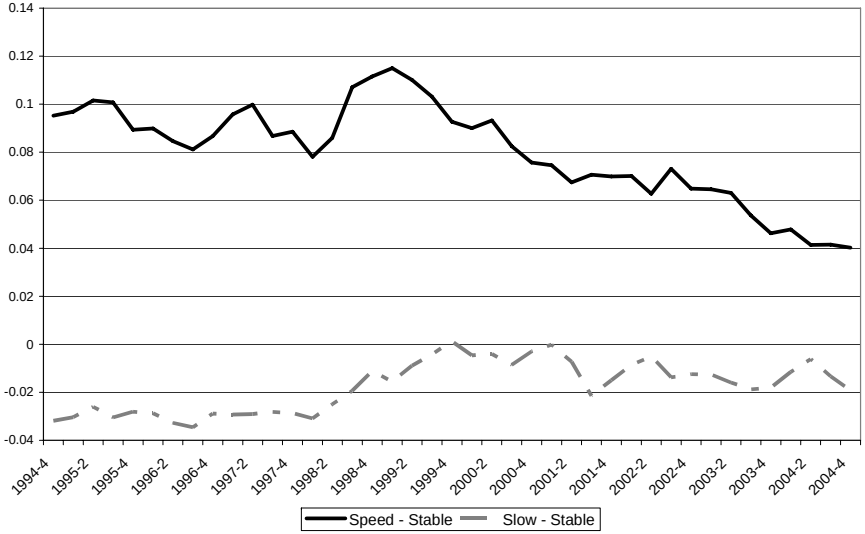


Figure 2F
Mean Difference in VOLUME

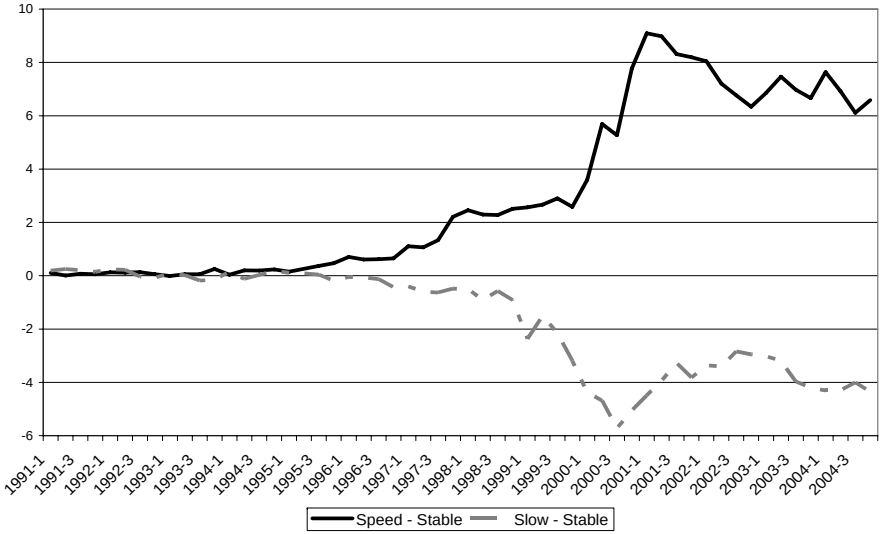


Figure 2G
Mean Difference in RATING

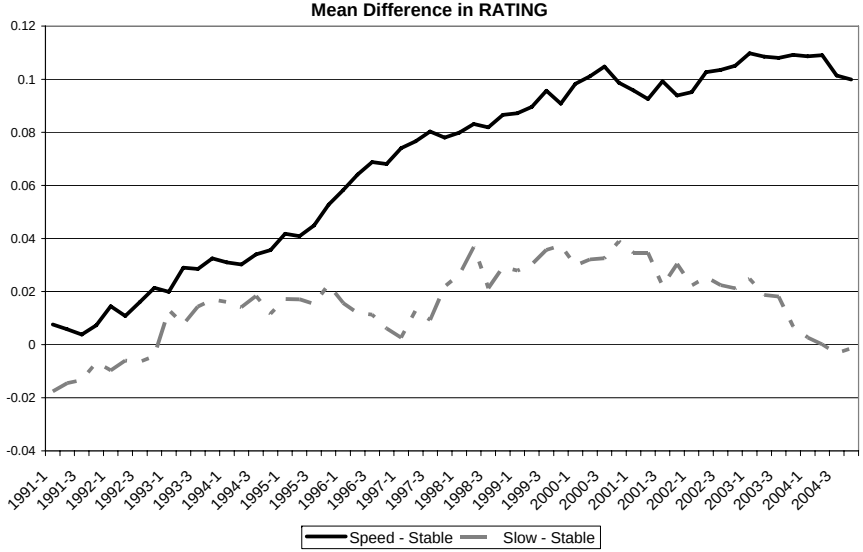
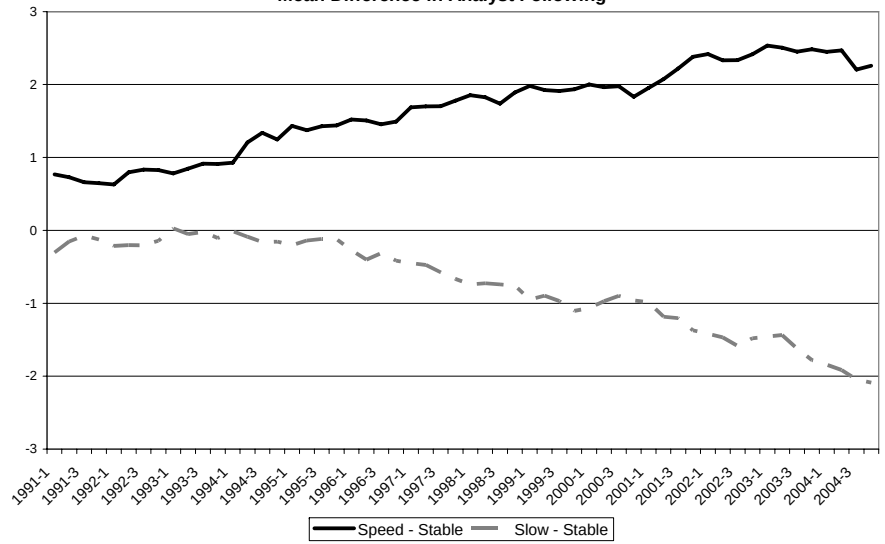


Figure 2H
Mean Difference in Analyst Following



The above figures compare speeding versus slowing firms in terms of data collection speed (Figure 2A), overall institutional ownership (Figure 2B), proportion of firms in major market indices (Figure 2C-2E), trading volume in millions of shares (Figure 2F), proportion of firms with S&P credit ratings (Figure 2G), and number of analyst following (Figure 2H). To identify firms whose collection lags have trended upwards versus downwards over time, we run the following regression for each firm:

$$BLOM\_DL_i = \alpha + \beta TIME_q + \varepsilon_i$$

where $BLOM\_DL_i$ is the rank of *DL*. To eliminate the effect of EDGAR on Compustat collection speed, we compute Blom normal scores (Blom 1958) from the ranks. $TIME_q$ is a time trend variable for calendar quarter over 1991Q2-2004Q4 (ranging from 1 to 55). We identify firms as having inter-temporally faster (slower) information dissemination speed if their β is significantly negative (positive) at the two-tailed level of 5 percent and denote them *SPEED* (*SLOW*). We define stable firms (denoted as *STABLE*) as those whose β estimate is insignificant at the two-tailed level of 5 percent. We use the stable firms as a benchmark and graph mean values of *SPEED* and SLOW groups as deviations from those of the *STABLE* group.

TABLE 1
Sample Selection and Descriptive Statistics on Computat Data Collection Lags

Panel A. Sample for analysis of periodic SEC filings collection speed

Firm-quarters in Compustat Preliminary History, Compustat AFR database, CRSP, and SEC filing date database over the period 1991-2004	329,048
Less:	
Firm-quarters with missing Compustat production date (FINALQPRD)	(4,664)
Filing date preceding fiscal quarter end	(49)
FINALQPRD preceding filing date	(1,594)
Firm-quarters with Compustat collection lag longer than 99 percentile (129 weekdays)	(3,217)
Final Sample	319,524

Panel B: Descriptive Statistics on S&P dissemination lag of Periodic SEC Filings (DL)

	N	Mean	Std	Min	Q1	Median	Q3	Max
All	319,524	18.437	15.562	0	8	15	23	129
Pre-EDGAR (1991-1995)[†]	97,741	24.696	16.147	0	16	22	29	129
Post-EDGAR (1997-2004)[†]	195,707	14.116	12.640	0	7	11	18	129

Panel A of this table describes our sample selection process which yields 319,524 firm-quarters in the final sample. However, the actual sample size varies with financial data availability in the tables reporting regression results. Panel B reports the summary statistics of Compustat collection lag of SEC 10-K/10-Q filings over 1991-2004 and separately for pre- and post-EDGAR periods. DL is measured as the number of weekdays from a company's 10-K/10-Q filing date to Compustat production date when the company's finalized financial data first appears in Compustat (FINALQPRD).

[†] Although EDGAR became effective on May 6, 1996, our discussions with S&P indicate that Compustat had not obtained the full benefits of EDGAR in the year of transition. So we eliminate year 1996 (26,076 firm-quarters) from both pre- and post-EDGAR periods.

TABLE 2
Descriptive Statistics

Panel A: Firm-level factors (N = number of firm-quarters)

	N	Mean	Std	Median	Min	Max	Q-range
$INST_{i,q-1}$	251,884	33.659	26.508	28.710	0	100	44.293
$TRA_{i,q-1}$	251,884	9.860	12.015	5.262	0	99.781	14.289
$QIX_{i,q-1}$	251,884	15.752	14.227	11.967	0	100	20.755
$DED_{i,q-1}$	251,884	7.150	9.517	3.266	0	96.824	11.295
$SPECTRUM_{i,q-1}$	319,524	0.788	0.409	1	0	1	0
$ARBRISK_{i,q}$	308,802	3.941	2.791	3.190	0	19.460	3.046
$PRICE_{i,q}$	317,170	16.211	16.806	11	0.0001	102	19.183
$VOLUME_{i,q}$	309,120	6.808	25.269	0.358	0	235.778	2.415
$S\&P500_{i,q}$	319,524	0.074	0.262	0	0	1	0
$CAP1B_{i,q}$	315,524	0.096	0.295	0	0	1	0
$S\&P1000_{i,q}$	315,524	0.084	0.277	0	0	1	0
$RATING_{i,q}$	319,494	0.195	0.396	0	0	1	0
$FINSERV_i$	319,524	0.187	0.390	0	0	1	0
$UTILITY_i$	319,524	0.028	0.166	0	0	1	0
$SPECIAL_{i,q}$	319,524	0.221	0.415	0	0	1	0
$EXTRAORD_{i,q}$	319,524	0.043	0.203	0	0	1	0
$CONCUR_{i,q}$	309,614	0.201	0.401	0	0	1	0

Panel B: Macro-level factors (N = number of Compustat production cycles)

	N	Mean	Std	Median	Min	Max	Q-range
$\#EA_p$	703	0.429	0.385	0.268	0.010	1.680	0.546
$\#FL_p$	703	0.451	0.781	0.188	0.010	4.481	0.268
$\#EA\_UPDATE_p$	703	0.429	0.384	0.268	0.012	2.050	0.538
$\#FL\_UPDATE_p$	703	0.451	0.367	0.341	0.022	2.425	0.430
EA_BKLG_{p-1} [†]	703	0.233	0.216	0.152	0.013	1.263	0.234
FL_BKLG_{p-1} [†]	703	1.713	1.307	1.223	0.120	7.069	1.732
$QTREND_p$	703	0.078	0.269	0	0	1	0
$YREND_p$	703	0.020	0.140	0	0	1	0

This table reports number of observations, mean, median, standard deviation, minimum, maximum, and inter-quartile range (Q-range) of explanatory variables in testing the determinants of S&P information dissemination lag. Panel A reports summary statistics of firm-level variables. All variables except the last three in Panel A proxy for demand-side forces for corporate financial information. The last three variables in Panel A (*SPECIAL*, *EXTRAORD* and *CONCUR*) and variables in Panel B, which reports summary statistics of macro-level variables, proxy for supply-side constraints. These variables are regressors in Table 3 and defined in Appendix.

[†] 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).

TABLE 3
Poisson Regression of Compustat Collection Lags on Institutional Demand and Other Determinants (N = 293,577 firm-quarters) [†]

		Regression 1			Regression 2		
	Pred. Sign	*Coeff.*	*p value[‡]*	*Marginal Effect*	*Coeff.*	*p value[‡]*	*Marginal Effect*
Demand-side factors:							
$INST_{i,q-1}$	-	-0.0010	0.000***	-4.58%			
$TRA_{i,q-1}$	?				0.0000	0.845	-0.05%
$QIX_{i,q-1}$	-				-0.0024	0.000***	-4.99%
$DED_{i,q-1}$	?				-0.0001	0.501	-0.12%
$SPECTRUM_{i,q-1}$	?	-0.0183	0.001***	-1.83%	-0.0147	0.005***	-1.47%
$ARBRISK_{i,q}$	+	0.0021	0.003***	0.63%	0.0018	0.009***	0.54%
$PRICE_{i,q}$	-	-0.0007	0.000***	-1.25%	-0.0006	0.000***	-1.20%
$VOLUME_{i,q}$	-	-0.0007	0.000***	-0.17%	-0.0008	0.000***	-0.18%
$S\&P500_{i,q}$	-	-0.4119	0.000***	-41.19%	-0.3941	0.000***	-39.41%
$CAP1B_{i,q}$	-	-0.1282	0.000***	-12.82%	-0.1251	0.000***	-12.51%
$S\&P1000_{i,q}$	-	-0.2239	0.000***	-22.39%	-0.2152	0.000***	-21.52%
$RATING_{i,q}$	-	-0.0458	0.000***	-4.58%	-0.0435	0.000***	-4.35%
$FINSERV_i$	+	0.0883	0.000***	8.83%	0.0907	0.000***	9.07%
$UTILITY_i$	+	0.2197	0.000***	21.97%	0.2218	0.000***	22.18%
Supply-side factors:							
$SPECIAL_{i,q}$	+	0.0173	0.000***	1.73%	0.0172	0.000***	1.72%
$EXTRAORD_{i,q}$	+	0.0701	0.000***	7.01%	0.0691	0.000***	6.91%
$CONCUR_{i,q}$		-0.0191	0.000***	-1.91%	-0.0196	0.000***	-1.96%
EA_BKLG_{p-1}		0.0406	0.019**	1.55%	0.0378	0.028**	1.44%
$(EA\_BKLG_{p-1})^2$		0.0117	0.259	0.36%	0.0133	0.199	0.41%
FL_BKLG_{p-1}		0.1831	0.000***	27.92%	0.1823	0.000***	27.80%
$(FL\_BKLG_{p-1})^2$		-0.0104	0.000***	-6.05%	-0.0102	0.000***	-5.95%
$\#EA_p$		0.3105	0.000***	3.82%	0.3244	0.000***	3.99%
$(\#EA_p)^2$		-0.9862	0.000***	-2.65%	-1.0221	0.000***	-2.75%
$\#FL_p$		0.1595	0.000***	17.99%	0.1587	0.000***	17.90%
$(\#FL_p)^2$		-0.0429	0.000***	-6.30%	-0.0426	0.000***	-6.27%

(Continued on next page)

[†] For brevity, we suppress the slope estimates of quarter-/year-end dummies and weekday dummies.
[‡] ***, **, * represent statistical significance at 0.01, 0.05, and 0.1 levels respectively.

TABLE 3 (cont'd)

		Regression 1			Regression 2		
	Pred. Sign	*Coeff.*	*p value*	*Marginal Effect*	*Coeff.*	*p value*	*Marginal Effect*
Year effects:							
Y1992		-0.1770	0.000***	-17.70%	-0.1758	0.000***	-17.58%
Y1993		-0.1196	0.000***	-11.96%	-0.1190	0.000***	-11.90%
Y1994		-0.0767	0.000***	-7.67%	-0.0761	0.000***	-7.61%
Y1995		0.0026	0.701	0.26%	0.0025	0.709	0.25%
Y1996		-0.1195	0.000***	-11.95%	-0.1227	0.000***	-12.27%
Y1997		-0.6013	0.000***	-60.13%	-0.6055	0.000***	-60.55%
Y1998		-0.3231	0.000***	-32.31%	-0.3183	0.000***	-31.83%
Y1999		-0.3030	0.000***	-30.30%	-0.3010	0.000***	-30.10%
Y2000		-0.6305	0.000***	-63.05%	-0.6277	0.000***	-62.77%
Y2001		-0.7019	0.000***	-70.19%	-0.7071	0.000***	-70.71%
Y2002		-0.9271	0.000***	-92.71%	-0.9339	0.000***	-93.39%
Y2003		-0.8665	0.000***	-86.65%	-0.8656	0.000***	-86.56%
Y2004		-0.7377	0.000***	-73.77%	-0.7309	0.000***	-73.09%
INTERCEPT		2.8166	0.000***		2.8178	0.000***	
Pseudo R^2			0.2275			0.2278	

This table reports the poisson regression results where S&P collection lag is the dependent variable:

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

When data is not available in Spectrum, we set the institutional ownership variables ($INST_{i, q-1}$, $TRA_{i, q-1}$, $QIX_{i, q-1}$, $DED_{i, q-1}$) to zero and include $SPECTRUM_{i, q-1}$ to control for lack of data availability from Spectrum. For each specification, we report estimated coefficients in the first column, followed by two-tailed p value (based on Huber-White standard errors) in the second column, and the marginal effect in the third column. The marginal effect is 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.

Table 4
Interaction Analysis Results of Institutional Ownership and Stock Characteristics
(N = 293,577 firm-quarters)

Panel A: Interaction of Institutional Ownership and Arbitrage Cost

	Low arbitrage cost	Medium arbitrage cost	High arbitrage cost	Low vs. High
$TRA_{i,q-1}$	-0.0002	0.0001	-0.0001	0.0001
	(0.548)	(0.649)	(0.872)	(0.869)
$QIX_{i,q-1}$	-0.0031	-0.0024	0.0001	0.0032
	(0.000)***	(0.000)***	(0.873)	(0.000)***
$DED_{i,q-1}$	0.0017	-0.0008	-0.0028	-0.0045
	(0.000)***	(0.005)***	(0.000)***	(0.000)***

Panel B: Interaction of Institutional Ownership and Membership in Major Market Indices

	S&P500	CAP1B	S&P1000	OTHER	S&P 500 vs. OTHER
$TRA_{i,q-1}$	-0.0009	-0.0008	-0.0016	0.0006	0.0015
	(0.337)	(0.100)*	(0.001)***	(0.008)***	(0.116)
$QIX_{i,q-1}$	-0.0058	-0.0047	-0.0039	-0.0009	0.0049
	(0.000)***	(0.000)***	(0.000)***	(0.000)***	(0.000)***
$DED_{i,q-1}$	0.0015	0.0015	0.0080	-0.0021	-0.0036
	(0.126)	(0.011)**	(0.000)***	(0.000)***	(0.000)***

This table reports slope estimates from poisson regressions of S&P collection lag on interactions of institutional ownership with investor types (Panel A) and with stock characteristics (Panel B). All other control variables in Table 3 are also included in the regressions, but their slope estimates are suppressed for brevity. Two-tailed p values (based on Huber-White standard errors) are provided in parentheses. In Panel A we interact institutional ownership with arbitrage risk and transaction cost. We first construct a combined measure for total arbitrage costs by ranking (from 0 to 2) *ARBRISK* (ascending), *PRICE* (descending), and *VOLUME* (descending) in terciles and add up the rankings to get a combined score. The final score ranges from 0 to 6. We then create three dummy variables indicating low arbitrage costs for scores 0 and 1, medium arbitrage costs for scores 2 to 4, and high arbitrage costs for scores 5 and 6, respectively. We interact our institutional ownership variables with these three dummy variables. In Panel B we interact institutional ownership with dummy variables indicating major market indices. The three of the dummy variables representing the major stock market indices (S&P500, CAP1B, and S&P1000) are defined in Appendix. The firm-years where none of the three dummy variables take a value of one are assigned to OTHER.

Table 5
Firm and Time Fixed Effects Regression of Changes in Institutional Ownership by Trading Type on S&P Data Collection Lag and Control Variables (N = 223,894 firm-quarters)[†, ‡]

	Pred. Sign	$\Delta TRA_{i,q}$	$\Delta QIX_{i,q}$	$\Delta DED_{i,q}$
$DL\_KNOWN_{i,q-1}$	-/-/?	-0.0162	-0.0176	0.0139
		(0.000)***	(0.000)***	(0.000)***
$INFOSTATUS_{i,q-1}$	-/-/?	-0.2590	-0.5737	0.3323
		(0.213)	(0.005)***	(0.001)***
$OWN_{i,q-1}$		-0.2615	-0.2237	-0.2204
		(0.000)***	(0.000)***	(0.000)***
$PORTWT_{i,q-1}$		-0.8282	-1.8016	-0.0563
		(0.001)***	(0.000)***	(0.035)**
$Ln(MVE)_{i,q-1}$		0.6695	1.0937	-0.0454
		(0.000)***	(0.000)***	(0.392)
$BTM_{i,q-1}$		-0.1726	0.4061	-0.2951
		(0.085)*	(0.006)***	(0.000)***
$RETQ24_{i,q}$		0.0025	-0.0001	0.0005
		(0.023)**	(0.918)	(0.231)
$RETQ1_{i,q}$		0.0134	-0.0009	0.0007
		(0.000)***	(0.614)	(0.237)
$RETQ0_{i,q}$		0.0086	-0.0003	0.0000
		(0.000)***	(0.776)	(0.999)
$UE_{i,q}$		0.6853	0.9926	-0.1163
		(0.000)***	(0.000)***	(0.350)
Adjusted R^2		0.1376	0.1063	0.1133

This table reports the results of firm and time fixed-effects regressions used in Ke and Petroni (2004):

$$\Delta OWN_{i,q} = \beta_0 + \beta_1 DL\_KNOWN_{i,q-1} + \beta_2 INFOSTATUS_{i,q-1} + \delta_3 OWN_{i,q-1} + \delta_4 PORTWT_{i,q-1} + \delta_5 Ln(MVE_{i,q-1}) + \delta_6 BTM_{i,q-1} + \delta_7 RETQ24_{i,q} + \delta_8 RETQ1_{i,q} + \delta_9 RETQ0_{i,q} + \delta_{10} UE_{i,q} + \varepsilon_{i,q}$$

All variables are as defined in Appendix.

[†] Two-tailed p-values are included in parentheses. The p values are based on two-way clustered standard errors that adjust for correlations within firms and within time periods.
[‡] ***, **, * represent statistical significance at 0.01, 0.05 and 0.1 levels respectively.

TABLE 6
Sensitivity Analyses Using the Number of Institutions as a Proxy for Information Demand[†, ‡]

Panel A: Replication of Table 3 (N = 293,577 firm-quarters)

	Pred. Sign	Regression 1	Regression 2
$NUMINST_{i,q-1}$	-	-0.0280	
		(0.000)***	
$NUMTRA_{i,q-1}$	?		-0.0127
			(0.000)***
$NUMQIX_{i,q-1}$	-		-0.0295
			(0.000)***
$NUMDED_{i,q-1}$	?		0.0279
			(0.000)***

Panel B: Replication of Table 5 (N = 223,894 firm-quarters)

	Pred. Sign	$\Delta NUMTRA_{i,q}$	$\Delta NUMQIX_{i,q}$	$\Delta NUMDED_{i,q}$
$DL\_KNOWN_{i,q-1}$	-/-/?	-0.0022	-0.0011	0.0021
		(0.000)***	(0.000)***	(0.000)***
$INFOSTATUS_{i,q-1}$	-/-/?	-0.0457	-0.0312	0.0495
		(0.007)***	(0.000)***	(0.005)***

This table reports the results of Table 3 and Table 5 regressions using the number of institutional investors by investor types as a proxy for demand for information. *NUMINST*, *NUMTRA*, *NUMQIX*, and *NUMDED* are natural logarithm of the number of all institutions, transients, quasi-indexers, and dedicated investors, respectively. All other variables are as defined in Appendix.

[†] Two-tailed p-values are included in parentheses. The p values are based on two-way clustered standard errors that adjust for correlations within firms and within time periods.
[‡] ***, **, * represent statistical significance at 0.01, 0.05 and 0.1 levels respectively.

TABLE 7
Sensitivity Analyses Excluding Bank Holdings from Institutional Ownership[†, ‡]

Panel A: Replication of Table 3(N = 293, 577 firm-quarters)

Selected Coefficients	*Pred. Sign*	Regression 1	Regression 2
$INST_{i,q-1}$	-	-0.0012	
		(0.000)***	
$TRA_{i,q-1}$	?		-0.0008
			(0.000)***
$QIX_{i,q-1}$	-		-0.0023
			(0.000)***
$DED_{i,q-1}$	?		-0.0007
			(0.002)***

Panel B: Replication of Table 5 (N = 223,894 firm-quarters)

Selected Coefficients	*Pred. Sign*	$\Delta TRA_{i,q}$	$\Delta QIX_{i,q}$	$\Delta DED_{i,q}$
$DL\_KNOWN_{i,q-1}$	-/-/?	-0.0159	-0.0204	0.0118
		(0.000)***	(0.000)***	(0.000)***
$INFOSTATUS_{i,q-1}$	-/-/?	-0.2633	-0.5647	0.2935
		(0.155)	(0.004)***	(0.003)***

This table reports the results of Table 3 and Table 5 regressions excluding the holdings of banks from the measures of institutional ownership. All variables are as defined in Appendix.

[†] Two-tailed p-values are included in parentheses. The p values are based on two-way clustered standard errors that adjust for correlations within firms and within time periods.
[‡] ***, **, * represent statistical significance at 0.01, 0.05 and 0.1 levels respectively.

APPENDIX
Variable Definitions

Variables used in Tables 3, 4, 6, and 7:

$DL_{iq,tp}$	= the number of weekdays from the 10-K/10-Q filing date t of firm i for fiscal quarter q to Compustat production date p (FINALQPRD) when the firm's finalized financial data first appeared in Compustat's Research Insight database.
$INST_{i,q-1}$	= percentage of institutional holdings of firm i (winsorized to 100) at the end of the calendar quarter immediately preceding the fiscal quarter q.
$TRA_{i,q-1}$	= percentage of stock ownership of firm i held by transient institutions at the end of the calendar quarter immediately preceding the fiscal quarter q (winsorized to 100).
$QIX_{i,q-1}$	= percentage of stock ownership of firm i held by quasi-indexers at the end of the calendar quarter immediately preceding the fiscal quarter q (winsorized to 100).
$DED_{i,q-1}$	= percentage of stock ownership of firm i held by dedicated institutions at the end of the calendar quarter immediately preceding the fiscal quarter q (winsorized to 100).
$NUMINST_{i,q-1}$	= natural logarithm of number of institutions of firm i at at the end of the calendar quarter immediately preceding the fiscal quarter q.
$NUMTRA_{i,q-1}$	= natural logarithm of number of transient institutions of firm i at at the end of the calendar quarter immediately preceding the fiscal quarter q.
$NUMQIX_{i,q-1}$	= natural logarithm of number of quasi-indexers of firm i at at the end of the calendar quarter immediately preceding the fiscal quarter q.
$NUMDED_{i,q-1}$	= natural logarithm of number of dedicated institutions of firm i at at the end of the calendar quarter immediately preceding the fiscal quarter q.
$SPECTRUM_{i,q-1}$	= 1 if institutional ownership data of firm i at the end of the calendar quarter immediately preceding the fiscal quarter q is available from Thomson Financial Spectrum database, and 0 otherwise.
$S\&P500_{i,q}$	= 1 if firm i is in the S&P 500 index at the end of fiscal quarter q, and 0 otherwise.
$CAP1B_{i,q}$	= 1 if $S\&P500_{i,q}$ equals 0 and the market capitalization of firm i is greater than one billion dollars at the end of fiscal quarter q, and 0 otherwise**.**
$S\&P1000_{i,q}$	= 1 if $CAP1B_{i,q}$ equals 0 and firm i is in the S&P MidCap 400 or SmallCap 600 index at the end of fiscal quarter q, 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 fiscal quarter q for firm i (a minimum of 5 observations is required). We multiply the standard deviation of residuals by 100 and winsorize it at its 99.5 percentile value.
$PRICE_{i,q}$	= closing stock price of firm i as of fiscal quarter end q, winsorized at its 99.5 percentile value.
$VOLUME_{i,q}$	= daily trading volume (in millions of dollars) of firm i averaged over fiscal quarter q, winsorized at its 99.5 percentile value**.**

$SPECIAL_{i,q}$	= 1 if a special item is reported in the Compustat database for firm i for fiscal quarter q, and 0 otherwise.
$EXTRAORD_{i,q}$	= 1 if an extraordinary item is reported in the Compustat database for firm i for fiscal quarter q, and 0 otherwise.
$RATING_{i,q}$	= 1 if S&P credit rating is available for firm i in fiscal quarter q, and 0 otherwise.
$FINSERV_i$	= 1 if financial service firms (SIC 6000-6999), and 0 otherwise.
$UTILITY_i$	= 1 if utilities (SIC 4900-4999), and 0 otherwise.
$CONCUR_{i,q}$	= 1 if the earnings announcement date of firm i matches its SEC periodic report filing date for fiscal quarter q, and 0 otherwise.
$\#EA\_UPDATE_p$	= number of preliminary earnings announcements (in '000s) from which data was collected during the Compustat production cycle ending at date p.
$\#FL\_UPDATE_p$	= number of periodic SEC reports (in '000s) from which data was collected during the Compustat production cycle ending at date p.
EA_BKLG_{p-1}	= number of uncollected earnings announcements (in '000s) at the end of Compustat production cycle p-1.
FL_BKLG_{p-1}	= number of uncollected 10-K/10-Q filings (in '000s) at the end of Compustat production cycle p-1.
$\#EA_p$	= number of earnings announcements (in '000s) made by all companies followed by Compustat during the production cycle ending at date p.
$\#FL_p$	= number of 10-K/10-Q filings (in '000s) made by all companies followed by Compustat during the production cycle ending at date p.
$YREND_p$	= 1 if the end of production cycle p is the last business day in a year, and 0 otherwise.
$QTREND_p$	= 1 if the end of production cycle p is the last business day in a quarter, and 0 otherwise.

Variables used in Table 5, 6, and 7:

Note: Variables measured for a fiscal quarter are based on information from the most recent fiscal quarter prior to the calendar quarter q.

$\Delta TRA_{i,q}$	= change in percentage of stock ownership of firm i held by transient institutions over the calendar quarter q.
$\Delta QIX_{i,q}$	= change in percentage of stock ownership of firm i held by quasi-indexers over the calendar quarter q.
$\Delta DED_{i,q}$	= change in percentage of stock ownership of firm i held by dedicated institutions over the calendar quarter q.
$INFOSTATUS_{i,q-1}$	= 1 if by the end of the institutional ownership measurement date the firm i has not yet filed its periodic SEC report or Standard & Poor's has not yet collected the information from the periodic report for the fiscal quarter, and 0 otherwise.
$DL\_KNOWN_{i,q-1}$	= $DL_{i,q-1}$ (zero) when $INFOSTATUS_{i,q-1}$ equals zero (one).
DL_{iq-1}	= the number of weekdays from the 10-K/10-Q filing date t of firm i for the fiscal quarter to Compustat production date when the firm's finalized financial data first appeared in Compustat's Research Insight database.

$OWN_{i,q-1}$	= $TRA_{i,q-1}$, $QIX_{i,q-1}$, or $DED_{i,q-1}$, when the dependent variable is $\Delta TRA_{i,q}$, $\Delta QIX_{i,q}$, or $\Delta DED_{i,q}$, respectively.
$PORTWT_{i,q-1}$	= mean portfolio weight of firm i's stock held by the institutions included in the calculation of the dependent variable (weighted by the total market capitalization of all stocks held by the same institutions), calculated as of the end of the calendar quarter q-1.
$Ln(MVE)_{i,q-1}$	= logarithm of market value of common equity of firm i at the beginning of fiscal quarter q.
$BTM_{i,q-1}$	= ratio of book value of equity to market value of equity of firm i at the beginning of fiscal quarter q.
$RETQ24_{i,q}$	= buy-and-hold raw return in percentage of firm i for two to four calendar quarters prior calendar quarter q.
$RETQ1_{i,q}$	= buy-and-hold raw return in percentage of firm i for the calendar quarter q.
$RETQ0_{i,q}$	= buy-and-hold raw return in percentage of firm i from 30 to 3 trading days before the earnings announcement date of fiscal quarter q.
$UE_{i,q}$	= change in earnings before extraordinary items of firm i from fiscal quarter q-4 to q.

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(3): 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 Earnings over Long-Run Value? *Contemporary Accounting Research* 18(2): 207-246.

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.

Dobrzynski, J., 1993. Relationship Investing. *Business Week*, March 15.

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

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.

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.

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. *Review of Financial Studies*, forthcoming.

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.

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.

A. Wang and G. Zhang, 2006. Institutional Equity Investment, Asymmetric Information and Credit Spreads. Working paper, University of California at Irvine and University of Missouri at St. Louis.

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.