Division of Research
School of Business Administration

January 1990

ANALYST FOLLOWING AND INSTITUTIONAL OWNERSHIP

Working Paper #629

Patricia C. O'Brien
The University of Michigan
and
Ravi Bhushan
Massachusetts Institute of Technology

The authors are grateful to Robert Egan and Melinda Su for research assistance, and to Lynch, Jones & Ryan for providing the analyst data. We received helpful comments on an earlier paper from seminar participants at Cornell, Duke, the University of Michigan, MIT and Washington University. Remaining errors and omissions are our responsibility.

Analyst Following and Institutional Ownership

Patricia C. O'Brien and Ravi Bhushan

1. Introduction

A growing literature on firms' information environments often cites analyst coverage as a component of this environment.[1] In this paper, we examine analysts' decisions to follow firms, along with institutional investors' decisions to hold these same firms in their portfolios. We view the two as interrelated through institutions' and brokers' customer/supplier relationship.

Analysts frequently perform or are associated with brokerage activities.[2] When analyst reports are used to market broker services to institutional customers or potential customers, the effect is to increase institutional ownership in followed firms. Likewise, institutional demand for information about particular firms is likely to affect analysts' decisions about which firms to follow. As investors, institutions have a straightforward demand for information as a basis for decisions. Standards of fiduciary responsibility for money managers may generate an additional link, as institutions can use analyst reports to support the prudence of their decisions.

If analysts' and institutions' behavior are jointly endogenously determined by many of the same firm and industry characteristics, single-equation models may contain simultaneous equations bias. We find different inferences from single-equation models and a simultaneous two-equation system, indicating this effect is present in the single-equation setting. In

particular, the association of firm size variables with analyst following may be artifacts of this bias, rather than a causal relation.

To focus on analysts' and institutions' concurrent decisions, we examine changes in analyst following and institutional ownership, rather than levels of these two variables. Our underlying motivation is similar to that of much previous work: to identify characteristics that influence firms' information environment. The difference in our focus is illustrated by the following example. A widely-noted firm characteristic related to the information environment is firm size. Bhushan (1989b) and Shores (1989), among others, note that large firms generally are more heavily followed by analysts. Our question is: if a firm grows, does its analyst following increase? Note that the cross-sectional empirical observation in previous work does not imply an affirmative answer to our time-series question.

Year-to-year changes in variables provide a stronger test of causal relations than levels of those variables, since the levels of many economic and non-economic variables are cross-sectionally correlated. Cross-sectional associations can often be found between variables that are not causally related.[3] While correlations in year-to-year changes do not imply causality either, a failure to find correlation in changes can help to distinguish between meaningful and spurious associations. That is, they provide a necessary, but not a sufficient, condition for a correctly-specified causal relation to exist.[4]

We investigate both firm and industry characteristics as determinants of analyst following and institutional ownership. Our results show that analyst following increases more in regulated industries and in growing industries; and that analyst following increases more in relatively neglected firms and in

firms whose volatility has declined. Institutional ownership increases with firm size, and with increased market risk.

The plan of the rest of the paper is as follows. In the next section we describe our methodology and rationales for the variables we suggest as determinants of analyst following and institutional ownership. Section 3 describes the data sources and some characteristics of the sample. In section 4 we present the results, and we conclude in Section 5.

2. Determinants of Analyst Following and Institutional Ownership

In this section we discuss variables we expect to be associated with changes in analyst following or changes in institutional ownership. Variants of some of them, such as firm size, have been used in previous studies and are known to be associated with firms' information environments.[5] We include industry characteristics that may influence analysts' decisions as well, to broaden the definition of environment.

We use analyst data from the I/B/E/S Summary tape produced by Lynch, Jones, & Ryan, covering the period January 1976 through June 1988. All stock return data and SIC codes are from the Center for Research in Securities Prices (CRSP) 1987 Daily Stock and NASDAQ files. Institutional ownership data are from Standard and Poor's (S&P) Security Owners' Stock Guide.

Since the number of analysts following any given firm tends to increase as the year end approaches, we collect the data near the fiscal year end, specifically in the eleventh month of the fiscal year.[6] For example: for a December year-end company, we use the November I/B/E/S Summary list, and the November issue of the S&P Stock Guide. We obtain CRSP data from roughly the same time: defining day 0 to be the year-end date, we obtain our stock-

related variables as of trading day -45, approximately two calendar months prior to the year end. This timing gives us roughly contemporaneous data from all three sources, allowing for a small time lag in the collection and publication of information in the I/B/E/S Summary and S&P Stock Guide.[7]

The variables are defined in Table 1. The joint endogenous (dependent) variables are the annual change in the number of analysts forecasting the firm's current year earnings per share (D#ANLST), and the annual change in the number of institutions holding the firm's common stock (D#INSTN).[8] These are proxies for the change in analyst following and the change in institutional ownership, respectively.

We expect analysts to prefer firms in industries with more sources of information. If information, for example about development of new technology or new products, is common to many firms in an industry, analysts may find scale economies in information collection. More firms in an industry mean more potential sources of information and a lower cost of information collection per firm. Thus, we expect industry growth to be associated with increased analyst following. We use 3-digit SIC codes to define industries, and define industry growth (SICGRO) to be the net change in the number of firms in the industry over the last 5 years, divided by the number of firms that were in the industry 5 years ago. We used all firms listed in the CRSP NYSE/AMEX and NASDAQ files to construct this variable, not just the firms that met our data requirements. Since the counts are restricted to listed firms, SICGRO will be inaccurate if any of the industries have substantial numbers of privately-held firms, and if these firms are followed by analysts.

We expect analysts to prefer regulated industries, because regulatory oversight and disclosures provide additional sources of operating information.

The industries we describe as regulated (REGIND = 1) are Trucking, Broadcasting, Utility Services, Savings Institutions, Securities Brokers, Insurance, Nursing and Personal Care, and Health.

The natural log of the market value of equity is often used as a proxy for firm size (LNMVEQ). We are interested in changes in firm size (DLNMVEQ), and use a simple decomposition to examine some possible causal links. Defining P_t and S_t as the price and number of shares outstanding at time t, respectively,

$$LNMVEQ_t = \ln (P_t * S_t) = \ln P_t + \ln S_t \quad , \qquad (1)$$

so:

$$DLNMVEQ_t = LNMVEQ_t - LNMVEQ_{t-1}$$

$$= (\ln [P_t / P_{t-1}]) + (\ln S_t - \ln S_{t-1}) \; . \qquad (2)$$

Equation (2) resolves the change in firm size into two parts, measuring the return on the firm's stock and changes in (log) shares outstanding, respectively.

If firm size, specifically the market value of a firm's equity, is important to analysts or institutions, then both parts of the right-hand-side of equation (2) should be associated with changes in analyst following and institutional ownership. However, depending on the reason for the association, these two components may differ. If analysts prefer to cover stocks that are performing well, and institutions prefer to buy these "winners," there will be a positive association between stock performance and changes in either analyst following or institutional ownership; but no link with changes in shares outstanding is implied. If a larger shareholder base means more business for analysts, this may create a positive association

between analyst changes and share increases, but implies nothing about the firm's stock return. Institutions may find it costly to own shares in companies with too few shares outstanding if they face regulatory restrictions on ownership or wish to avoid the SEC reporting requirements of large block holders. This implies institutional ownership may increase with increases in shares outstanding, but implies nothing about stock performance. These stories all suggest non-negative associations.

We use the change in the natural logarithm of the number of common shares outstanding on trading day -45 as a proxy for the change in the investor base (DLNSHRS). The proxy for stock performance is the market-adjusted return (MKTARET), the difference between the continuously compounded return on the stock and the return on the value-weighted index of all NYSE, AMEX and NASDAQ stocks[9] over a 250-trading-day period ending on trading day -45. MKTARET differs from the log of the relative price change in equation (2) by the dividend payout and the market adjustment. However, though it differs from the precise decomposition of the size variable, it provides a measure of firm-specific performance more in line with the rationales for its association with analysts' and institutions' decisions.

We expect volatility to be positively related to analyst following. Ceteris paribus, an increase in return variability increases the probability of a large expected return from private information, so trading profits for informed investors are likely to be higher for more volatile firms. Our proxy for volatility is the standard error of excess stock returns (RESIDSE), computed as the residual standard error from a market-model regression using the value-weighted index over the interval [-244,-45].[10] We use changes in volatility (DRESIDSE) as our regressor.

If institutional investors are more averse to negative outcomes than are other clienteles of investors, they may avoid firms with higher systematic risk, because adverse economic times will have more impact on these firms. We can imagine an asymmetric loss function arising, for example, because the risk of lawsuits is greater for institutions who manage money in trust for others than for other investors. This explanation indicates a negative relation between changes in systematic risk (DBETA) and changes in institutional ownership, ceteris paribus. We estimate systematic risk using the market model described in the preceding paragraph.

We expect preexisting conditions to affect analysts' and institutions' behavior. Specifically, we expect that other things equal, analysts will prefer firms with less competition from other analysts, and with more institutional owners. We expect that institutions will prefer more closely-followed firms and firms with more institutional owners, since both could be used to justify prudent business judgement. Our proxies for preexisting analyst following and institutional ownership are the prior-year levels of each (L#ANLST and L#INSTN, respectively).

In summary, based on the stories above we expect positive associations between changes in analyst following and changes in institutional ownership, industry growth, regulation, stock performance, changes in shares outstanding, changes in volatility, and lagged institutional ownership. We expect a negative association between changes in analyst following and lagged analyst following. Changes in institutional ownership are expected to be positively associated with changes in analyst following, stock performance, changes in shares outstanding, and lagged analyst following and institutional ownership.

We expect a negative relation between changes in institutional ownership and changes in systematic risk.

If analysts' decisions to cover a firm and institutions' decisions to hold the stock are jointly endogenously determined, then least-squares coefficients may be biased and inconsistent. Simultaneity problems will arise here if the unexplained portions of analyst following and institutional ownership are correlated. This induces correlation between the regressors and the error term in a single-equation model, a violation of OLS assumptions. The usual solution to eliminate the simultaneity problem is to identify the system, either by specifying regressors which influence one but not the other endogenous variable or by placing restrictions on the cross-equation covariances, and estimate the model using simultaneous equations techniques.

We identify the system by excluding one exogenous variable from each equation. The argument for BETA's influence was that institutions form a clientele with distinct preferences about systematic risk. DBETA should influence analyst following only through its effect on institutions, and is therefore a natural candidate for exclusion from the analyst equation. The argument for SICGRO was based on analysts' costs of becoming informed about industry production processes. It therefore may affect institutional ownership only via its effect on analyst following, so we exclude it from the institutions equation to just-identify the system.

We estimate the just-identified two-equation system using two-stage least squares (2SLS). For comparison, we also estimate each equation using ordinary least squares (OLS).

Anticipating substantial amounts of exploratory data analysis to specify the regression relation, we used the subset of data from fiscal years 1985

through 1987 for exploratory analysis and model specification. This leaves us with a hold-out sample of earlier fiscal years to test whether the inferences generalize to a different time period, reducing the chance that we have overfit our model to our specific sample. The results, reported in section 4 following our description of the sample selection in section 3, are reasonably robust.

3. Sample Selection and Data Description

Table 2 contains a description of the initial sample selection. Of 5811 firms listed in the I/B/E/S file, 4920 are in the combined CRSP NYSE/AMEX and NASDAQ files. Since we use industry characteristics in our analysis, we omit 666 firms with no SIC code listed on CRSP, or that changed SIC codes between January 1976 and December 1986.[11] The remaining 4254 firms represent 318 3-digit SIC industries. We restrict our analysis to industries with at least five firms, to ensure sufficient data to define the industry characteristics. This subset of 155 industries contains 3887 firms.

To reduce the burden of hand-collecting data, we select approximately one quarter of the 155 industries for the final sample. We select every fourth SIC code in numerical order, leaving 38 industries with 1104 firms. Since SIC codes are numerically related (e.g., 3-digit codes are subsets of 2-digit codes), selecting every fourth industry gives us a broad-based set of industries.

Of the 1104 firms, 846 have complete data in any year from all sources, including the S&P Security Owners' Stock Guide. The requirement of complete data in consecutive years for first-differencing reduces the sample to 716 firms with 3000 firm-years. Our two subsets, from fiscal years 1985 through

1987 and from fiscal years 1981 through 1984 have 1347 and 1653 firm-years, respectively.[12]

Tables 3 and 4 give information on the industry and time composition of the sample. In Table 3, the 38 industries are reported by name, along with the number of sample firms in each. SIC 671, Holding Offices, contains 31% of the sample firms. It is not a clearly defined industry, in which the member firms are expected to share operating characteristics or have intra-industry information transfers. However, its firms are legitimate observations in other, firm-specific respects. We perform the analysis both with and without SIC 671. Results are weaker, but do not contradict those reported here, when Holding Offices are excluded.

Panel A of Table 4 reports the distribution of observations by calendar year and fiscal year end. As expected, our sample is dominated by December year ends, but approximately one-quarter of the observations are for other year ends. Observations are roughly evenly distributed across years. Panel B of Table 4 reports some characteristics of the sample distributions of the variables over all 3000 firm-year observations.

In Table 5 we report pairwise correlations between variables, for the full sample and each of the subsamples. These are fairly stable between sub-periods. Although we indicate statistical significance at the 5% level, we do not wish to draw inferences from these correlations, first because we are interested in the multiple correlations, and second because statistical significance in the pairwise calculations is based on the assumption that all observations are independent draws. This is clearly violated for some of the exogenous and predetermined variables. For example, the lagged number of institutions holding a firm will be correlated from year to year.

The correlations in Table 5 are important for comparing inferences from the annual changes we examine here with the levels-form correlations reported in other studies. If we find no association between changes in the exogenous variables and changes in analyst following or institutional ownership, it may be because (1) the levels-form associations do not describe causal relations, or (2) we have misspecified the association, e.g. by failing to adjust for lagged responses. In Table 5, we find correlations between changes in analyst following and institutional ownership and changes in firm and industry characteristics. Moreover, we did not find stronger associations in the leads and lags we investigated.

4. Results

As discussed above, we use a simultaneous equations approach to model analyst following and institutional ownership jointly, and perform our initial estimates using the 1985 through 1987 subperiod. In Table 6, we present results for single equation regressions as well as simultaneous estimation of the following two-equation system.

$$\begin{aligned} D\#ANLST_{it} = {} & a_0 + a_1\, D\#INSTN_{it} + a_2\, SICGRO_{it} + a_3\, REGIND_{it} \\ & + a_4\, MKTARET_{it} + a_5\, DLNSHRS_{it} \\ & + a_7\, DRESIDSE_{it} + a_8\, L\#ANLST_{it} + a_9\, L\#INSTN_{it} + e_{it} \end{aligned} \tag{3}$$

and

$$\begin{aligned} D\#INSTN_{it} = {} & b_0 + b_1\, D\#ANLST_{it} + b_3\, REGIND_{it} \\ & + b_4\, MKTARET_{it} + b_5\, DLNSHRS_{it} + b_6\, DBETA_{it} \\ & + b_7\, DRESIDSE_{it} + b_8\, L\#ANLST_{it} + b_9\, L\#INSTN_{it} + u_{it} \end{aligned} \tag{4}$$

Within each subsample, we pool time-series and cross-section observations. Since the data are annual changes, the risk of pooling correlated observations is not as great as when variables are measured in levels, but this still may be a concern. For example, if changes in analyst following or institutional ownership are serially correlated, then consecutive years' observations for a single firm are not independent draws. We tested for this possibility by sorting the data by year within firm, and computing a Durbin-Watson statistic for first-order serial correlation in the residuals. In all instances the statistic was near 2, and failed to reject the hypothesis of no autocorrelation.

Two single-equation OLS regressions are reported for each endogenous variable in Table 6, the first including all possible regressors, and the second in the same form as the just-identified 2SLS model of equations (3) and (4). The results reported in the 2SLS column for each endogenous variable are from a joint estimation of the two equation system. The OLS regressions explain about 13% of the variation in changes in analyst following, and about 41% of the variation in changes in institutional ownership. The 2SLS estimation explains about 8% in the analyst regression, and about 39% in the institutions regression.

From the 2SLS results, it appears that changes in analyst following are positively associated with industry growth and regulation, and negatively associated with preexisting analyst following, as expected from our discussion above. Our expectation that analysts would prefer volatility is refuted by the significantly negative coefficient on that variable. Changes in institutional ownership are positively associated with stock performance, increases in shares outstanding, lagged analyst following and lagged

institutional ownership, as expected. Institutions do not appear to shun systematic risk, as we suggested, since DBETA has a significant positive coefficient.

If the errors from the OLS analyst and institutions regressions are correlated, then the coefficients are biased and inconsistent. Evidence of this is seen by comparing the OLS and 2SLS results. In the analyst equation, the intercept and the coefficients on industry growth, regulation and volatility changes are virtually unaffected by the 2SLS estimation. The statistical significance of lagged analyst following is reduced somewhat, but the coefficient changes little. The variables most affected are stock performance (MKTARET), changes in institutional ownership (D#INSTN) and lagged institutional ownership. All three coefficients are lower under 2SLS, and go from strong statistical significance to none. In the institutions regression, systematic risk is virtually unaffected. The statistical significance of stock performance, change in shares outstanding, lagged analyst following and lagged institutional ownership are reduced, although the coefficients remain similar and are still significantly different from 0. The greatest effect is on the joint endogenous variable, which is significantly related to institutional ownership in the OLS regressions, but not in the 2SLS estimation.

Taken together, we conclude from these results that institutions are likely to add growing firms to their portfolios. However, once we have accounted for the feedback effects between analysts' and institutions' behavior, we find no evidence that analysts move to cover growing firms. This result does not negate the fact that analyst following and firm size are related, but refutes a behavioral or causal link between the two.

We test the robustness of our results by performing the same estimation on a the earlier subperiod, with observations from fiscal 1981 through 1984. These results are reported in Table 7. As expected in holdout sample tests, the R^2 is lower for these regressions than in their counterparts in Table 6, ranging from 5% to 10% in the analyst equation and from 10% to 14% in the institutions equation. The tenor of most of the results reported above is preserved, however.

The 2SLS results show changes in analyst following positively related to industry growth and regulation, and negatively related to volatility and lagged analyst following. As before, the association between stock performance and analyst following apparent in the OLS results disappears in the 2SLS regression. The only major difference between the results on analyst following in the 1985-87 period and the 1981-84 period is that changes in analyst following are strongly positively related to lagged institutional ownership in the earlier subperiod. This was the relation we expected, but it was not evident in the original results.

Institutional ownership, as before, is positively associated with stock performance, changes in shares outstanding, changes in systematic risk, and lagged analyst following. The major difference in results in this earlier subperiod is in the coefficient on lagged institutional holding, which is zero instead of positive as expected.

5. Conclusions

To summarize, we find evidence that supports a behavioral link between analysts decisions to follow firms and industry characteristics that may indicate more sources of available information: growth and regulation. We

find that analysts tend to avoid volatility and competition from preexisting analyst following. We find no evidence of a causal link between changes in analyst following and changes in firm size, once the simultaneity with institutions' decisions is eliminated. Our evidence supports a behavioral association between institutions' decisions to hold firms' common stock and changes in firm size (both price changes and changes in shares outstanding), and prior analyst following. Somewhat paradoxically, we find that institutions seem to prefer increases in risk. Overall, our results are reasonably robust when the model is tested in an earlier time period.

Since the systematic risk results are opposite those we anticipated in our a priori arguments, and are maintained in the simultaneous equation estimation and the holdout sample, they bear some discussion. The positive relation could result from agency problems between institutional fund managers and the individuals whose money they manage. For example, if the managers are compensated for achieving high return, and the systematic risk of the portfolio is not monitored, then they may have incentive to invest in high-risk securities to increase the portfolio's expected return. This is the opposite of our a priori story, in which the fund managers' fiduciary responsibility made them more risk-averse than most investors.

We examine analyst following from a somewhat different perspective than most prior studies of the information environment. By examining annual changes in this variable along with changes in institutional ownership, we produce stronger tests of the causal relations between firm and industry characteristics and analysts' and institutions' decisions. The lack of evidence that increasing firm size will increase analyst following, once feedback effects between institutions and analysts are considered, calls into

question some conventional assumptions about how size affects information production. If analyst following is a reasonable proxy for information production, then our results fail to support a direct causal link between it and size.

REFERENCES

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

Badrinath, S.G., G.D. Gay, and J.R. Kale, 1988, Patterns of institutional investment and the managerial "safety-net" hypothesis, Working paper (Northeastern University and Georgia State University).

Bhushan, R., 1989a, Collection of information about publicly traded firms: theory and evidence, Journal of Accounting and Economics 2/3, 183-206.

Bhushan, R., 1989b, Firm characteristics and analyst following, Journal of Accounting and Economics 2/3, 255-274.

Brown, L., G. Richardson, and S. Schwager, 1987, An information interpretation of financial analyst superiority in forecasting earnings, Journal of Accounting Research 25, 49-67.

Collins, D.W., S.P. Kothari, and J.D. Rayburn, 1987, Firm size and information content of prices with respect to earnings, Journal of Accounting and Economics 9, 111-138.

Freeman, R., 1987, The association between accounting earnings and security returns for large and small firms, Journal of Accounting and Economics 9, 195-228.

Grant, E., 1980, Market implications of differential amounts of interim information, Journal of Accounting Research 18, 255-268.

Shores, D., 1989, The Association Between Interim Information and Security Returns Surrounding Earnings Announcements, working paper, University of Washington (January).

Skinner, D.J., 1989, Options Markets and the Information Content of Accounting Earnings Releases, working paper, University of Michigan (November).

1. See, for example, Shores (1989), Skinner (1989), Bhushan (1989a), and Collins, Kothari and Rayburn (1987).

2. This is particularly true for the I/B/E/S Summary database, which we use. I/B/E/S sells analyst forecasts to institutional investors, and includes only sell-side analysts, those associated with brokers, in the Summary forecast database.

3. For example, it is quite likely that the number of shares outstanding is cross-sectionally related to the number of parking-lot spaces owned by the firm, simply because firms differ in size and both variables tend to increase with size. However, it is far-fetched to think that a change in either of these variables would cause a change in the other.

4. Measurement problems and leads or lags in relations between variables may hinder attempts to find correlations empirically. We refer here to "correctly-specified" causal links to abstract from these empirical considerations until later in the paper.

5. Studies that suggest more information is generated for larger firms include Grant (1980), Atiase (1985), Collins, Kothari, and Rayburn (1987), Freeman (1987), and Bhushan (1989a).

6. Our fiscal year convention is illustrated by the following example: years ending between June 1985 and May 1986 are regarded as fiscal year 1985.

7. The I/B/E/S database is updated continuously as information arrives at Lynch, Jones & Ryan, and the Summary is produced in the third week of each month. The S&P Security Owners' Stock Guide states that each monthly issue is "revised through the last business day of the prior month." Therefore, data in the November issue are current as of the end of October.

8. The Security Owners' Stock Guide includes "nearly 2700" institutions "including investment companies, banks, insurance companies, college endowments, and 13-F money managers." S&P obtain the data from Vickers Stock Research Corp.

9. Preliminary results did not differ when an equally-weighted index from all markets was used. Results differed when only the NYSE/AMEX index was used.

10. We require at least 50 return observations in this interval to perform the estimation.

11. We began the study with the intention of collecting data for the period 1976-86. We added 1987 when the CRSP data for this year became available. We later found that the S&P Security Owners' Stock Guide began including pension funds in its definition of institutions during 1981, so we chose to work with the shorter span of years with consistent data.

12. The reported final sample numbers are after deletion of two outliers that influenced the regression results. In 1984, Energy Management Co. (CUSIP 29270110) had a measured change in volatility of 13.84, more than twice the range in volatility of all other observations. In 1985, Medical Care Int. (CUSIP 58450510) was the only sample firm in the Health and Allied Services

Industry (SIC 809), which had a 5-year growth rate of 13, more than quadruple the next highest observation. Each of these observations influenced one regression coefficient substantially (for volatility and industry growth, respectively), but had little effect on other coefficients.

TABLE 1

Definitions of Variables

Variable Name	Definition
#ANLST*	The number of analysts with estimates of current-year EPS listed in the I/B/E/S Summary file in the month prior to the fiscal year end.
#INSTN*	The number of institutions holding the stock, as reported in the S&P Security Owners' Stock Guide published in the month prior to the fiscal year end.
SIC	The three-digit SIC code listed in CRSP for the firm.
SICGRO	The net growth of the industry over the five-year period ending in the month prior to the fiscal year end. Net growth is the number of new firms listed in CRSP for the industry minus the number of firms exiting the industry, divided by the number of firms that were in the industry at the beginning of the period.
REGIND	Regulated industry: takes the value 1 if the industry is in SIC 421, 483, 493, 612, 621, 633, 805, or 809; and 0 otherwise. Industry names are in Table 3.
LNMVEQ*	The natural logarithm of the market value of equity (in $1000s) on trading day -45. Day 0 is the fiscal year end date. Data are from the CRSP files.
MKTARET	Market-adjusted return: the firm's continuously compounded return over trading days -294 to -45 minus the value-weighted market return over the same period (x 100). Day 0 is the fiscal year end date.
LNSHRS*	The natural logarithm of the number of shares outstanding, as listed in the CRSP files (cross-checked against data from S&P Guide).
BETA*	The estimated systematic risk of the stock, from a market model regression using a value-weighted market index, estimated over trading days -244 to -45. Day 0 is the fiscal year end date.
RESIDSE*	The residual standard error (x 100) from the market model regression described above in the definition of BETA.

*Variable names marked with an asterisk are used in first-differenced form. The regressors include lagged, undifferenced #ANLST and #INSTN.

TABLE 2

Sample Selection

Panel A: Intersection of Analyst and Stock Market Datasets

	I/B/E/S	NYSE/AMEX	NASDAQ
Firms in dataset	5811	5019	7659
Firms matched by CUSIP[1]	4920	2545	2854
Matched firms with no industry change[1]	4254	2267	2466

Panel B: Selection of Industries

	# firms	# industries
Intersection, from Panel A	4254	318
Industries with ≥ 5 firms	3887	155
Every 4th industry selected	1104	38

Panel C: Data Availability

	# firms	# industries	# firm-years
Data from all sources	846	38	3977
First-differenced data	716	38	3000
Subset: Fiscal 1985-87	570	38	1347
Subset: Fiscal 1981-84	664	38	1653

[1]The number of firms matched on I/B/E/S does not equal the sum of those matched on the NYSE/AMEX and those matched on NASDAQ, because some firms changed from OTC to one of the major exchanges (or the reverse) between 1976 and 1986, and so appear on both. We used CRSP's IPERM identification number to verify that these were identical firms.

TABLE 3

Sample Industries

SIC	Industry Name	# of sample firms
131	Crude Petroleum and Natural Gas	61
201	Meat Products	6
205	Bakery Products	3
221	Broadwoven Fabric Mills, Cotton	8
232	Men's and Boys' Furnishings	14
243	Millwork, Plywood & Structural Members	7
264	Misc. Converted Paper Products	11
275	Commercial Printing	15
283	Drugs	46
287	Agricultural Chemicals	4
301	Tires and Inner Tubes	9
324	Cement, Hydraulic	5
332	Iron and Steel Foundries	8
343	Plumbing and Heating, Except Electric	7
349	Misc. Fabricated Metal Products	18
354	Metalworking Machinery	13
358	Refrigeration and Service Machinery	11
364	Electrical Lighting and Wiring Equipment	14
369	Misc. Electrical Equipment and Supplies	12
379	Misc. Transportation Equipment	5
384	Medical Instruments and Supplies	40
394	Toys and Sporting Goods	9
421	Trucking and Couriers Services, Except Air	17
483	Radio and Television Broadcasting	13
493	Combination Utility Services	32
503	Lumber and Construction Materials	3
512	Drugs, Proprietaries and Sundries	10
533	Variety Stores	6
566	Shoe Stores	5
591	Drug Stores and Proprietary Stores	14
612	Savings Institutions	21
621	Security Brokers and Dealers	11
633	Fire, Marine and Casualty Insurance	15
671	Holding Offices	224
721	Laundry, Cleaning and Garment Service	4
751	Automotive Rentals, No Drivers	4
805	Nursing and Personal Care Facilities	8
809	Health and Allied Services, NEC	3
		716

TABLE 4

Sample Distribution Information

Panel A: Number of Observations by Fiscal Year End and Fiscal Year

Fiscal Year End	Number of Obs.	Fiscal Year[1]	Number of Obs.
1	79	81	338
2	41	82	411
3	64	83	431
4	44	84	473
5	41	85	499
6	115	86	455
7	35	87	393
8	71		3000
9	124		
10	71		
11	59		
12	2256		
	3000		

Panel B: Summary Statistics on Variables' Sample Distributions[2]

Variable Name	Mean	Std.Dev.	Median	Minimum	Maximum
D#ANLST	0.71	2.07	0.00	-9.00	13.00
D#INSTN	9.29	21.09	4.00	-85.00	202.00
SICGRO	0.14	0.33	0.13	-0.57	2.75
MKTARET	-0.03	0.34	0.00	-2.12	1.29
DLNSHRS	0.13	0.27	0.01	-2.05	2.52
DBETA	-0.01	0.39	0.00	-1.68	1.97
DRESIDSE	0.03	0.57	0.00	-2.33	4.21
L#ANLST	8.17	7.66	5.00	1.00	39.00
L#INSTN	81.84	117.61	38.00	1.00	771.00

[1]Our fiscal year convention is illustrated by the following example: years ending between June 1985 and May 1986 are regarded as fiscal year 1985.

[2]The summary statistics are computed over all 3000 firm-years in the sample.

TABLE 5

Pairwise Correlations of Variables[1]

Panel A: Full Sample (N=3000)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) D#ANLST	1.00	0.27*	0.05*	0.17*	0.11*	0.06*	-0.07*	0.07*	0.13*
(2) D#INSTN		1.00	-0.01	0.23*	0.26*	0.12*	0.01	0.35*	0.33*
(3) SICGRO			1.00	-0.11*	0.02	-0.07*	-0.11*	-0.01	0.01
(4) MKTARET				1.00	0.16*	0.14*	-0.20*	0.03	0.04*
(5) DLNSHRS					1.00	0.17*	0.10*	-0.01	-0.02
(6) DBETA						1.00	0.21*	0.03	0.03
(7) DRESIDSE							1.00	0.01	0.00
(8) L#ANLST								1.00	0.82*
(9) L#INSTN									1.00

Panel B: 1985-87 Subsample (N=1347)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) D#ANLST	1.00	0.31*	0.08*	0.24*	0.12*	0.05	-0.08*	0.05	0.10*
(2) D#INSTN		1.00	0.01	0.31*	0.28*	0.11*	-0.03	0.48*	0.49*
(3) SICGRO			1.00	0.05	0.03	-0.05*	-0.03	-0.03	-0.00
(4) MKTARET				1.00	0.23*	0.12*	-0.26*	0.09*	0.10*
(5) DLNSHRS					1.00	0.18*	0.07*	0.03	0.01
(6) DBETA						1.00	0.19*	-0.02	-0.02
(7) DRESIDSE							1.00	-0.04	-0.02
(8) L#ANLST								1.00	0.83*
(9) L#INSTN									1.00

Panel C: 1981-84 Subsample (N=1653)

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
(1) D#ANLST	1.00	0.25*	0.02	0.11*	0.10*	0.07*	-0.06*	0.11*	0.17*
(2) D#INSTN		1.00	-0.05	0.20*	0.23*	0.13*	0.00	0.12*	0.10*
(3) SICGRO			1.00	-0.25*	0.02	-0.08*	0.01	0.00	0.02
(4) MKTARET				1.00	0.11*	0.16*	-0.14*	0.00	0.00
(5) DLNSHRS					1.00	0.17*	0.10*	-0.06*	-0.06*
(6) DBETA						1.00	0.24*	0.10*	0.08*
(7) DRESIDSE							1.00	0.03	0.00
(8) L#ANLST								1.00	0.80*
(9) L#INSTN									1.00

---

[1] Pearson (product moment) correlations are computed across all firm-year observations. An asterisk (*) indicates that the correlation is significantly different from 0 at the 5% level or better, assuming the independent observations.

TABLE 6

OLS and 2SLS Regressions of Analyst Following and Institutional Investor Ownership on Firm and Industry Characteristics, 1985-87

Coefficient Estimates (t-statistics in parentheses):

	D#ANLST			D#INSTN		
	OLS	OLS	2SLS	OLS	OLS	2SLS
INTERCEPT	0.56 (5.60)	0.56 (5.59)	0.54 (4.95)	-0.28 (-0.31)	-0.29 (-0.33)	-0.14 (-0.05)
D#INSTN	0.03 (9.49)	0.03 (9.56)	0.05 (1.15)			
D#ANLST				2.21 (9.49)	2.21 (9.51)	1.98 (0.46)
SICGRO	0.33 (1.91)	0.32 (1.89)	0.31 (1.80)	-0.08 (-0.05)		
REGIND	0.35 (2.25)	0.35 (2.24)	0.36 (2.24)	-1.09 (-0.79)	-1.11 (-0.81)	-1.01 (-0.46)
MKTARET	0.90 (4.61)	0.91 (4.70)	0.60 (0.86)	12.48 (7.34)	12.48 (7.34)	12.79 (2.09)
DLNSHRS	0.08 (0.36)	0.09 (0.41)	-0.27 (-0.33)	17.01 (8.88)	17.01 (8.89)	17.15 (5.17)
DBETA	0.07 (0.47)			3.17 (2.47)	3.18 (2.48)	3.21 (2.22)
DRESIDSE	-0.18 (-1.68)	-0.17 (-1.61)	-0.18 (-1.65)	0.73 (0.79)	0.73 (0.79)	0.69 (0.58)
L#ANLST	-0.06 (-4.34)	-0.06 (-4.35)	-0.07 (-2.23)	0.76 (6.75)	0.76 (6.78)	0.75 (3.72)
L#INSTN[1]	0.20 (2.26)	0.19 (2.25)	0.09 (0.39)	4.59 (6.09)	4.59 (6.10)	4.67 (2.70)
Regression Summary Statistics						
# Observations	1347	1347	1347	1347	1347	1347
Adjusted R^2	0.1327	0.1332	0.0831	0.4140	0.4144	0.3939
Regression F	23.88	26.86	15.15	106.65	120.07	108.71
p-value for F	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

[1]The lagged number of institutions holding the firm is divided by 100 for this estimation, so the coefficient can be reported in two decimal places.

TABLE 7

OLS and 2SLS Regressions of Analyst Following and Institutional Investor Ownership on Firm and Industry Characteristics, 1981-84

Coefficient Estimates (t-statistics in parentheses):

	D#ANLST			D#INSTN		
	OLS	OLS	2SLS	OLS	OLS	2SLS
INTERCEPT	0.39	0.38	0.29	0.87	0.75	2.13
	(5.29)	(5.19)	(2.48)	(1.31)	(1.17)	(1.03)
D#INSTN	0.02	0.02	0.09			
	(8.71)	(8.80)	(1.60)			
D#ANLST				1.85	1.84	-1.08
				(8.71)	(8.68)	(-0.26)
SICGRO	0.32	0.32	0.34	-0.93		
	(2.31)	(2.25)	(2.13)	(-0.75)		
REGIND	0.34	0.35	0.36	-0.91	-0.81	0.06
	(2.80)	(2.87)	(2.58)	(-0.85)	(-0.76)	(0.04)
MKTARET	0.32	0.35	-0.19	7.37	7.59	8.89
	(2.31)	(2.55)	(-0.39)	(6.08)	(6.44)	(3.99)
DLNSHRS	0.31	0.34	-0.54	12.73	12.67	14.63
	(1.83)	(1.99)	(-0.69)	(8.56)	(8.53)	(4.60)
DBETA	0.17			2.28	2.32	2.96
	(1.32)			(2.05)	(2.08)	(1.99)
DRESIDSE	-0.22	-0.19	-0.19	-0.11	-0.11	-0.79
	(-2.58)	(-2.32)	(-1.96)	(-0.15)	(-0.15)	(-0.63)
L#ANLST	-0.03	-0.03	-0.05	0.33	0.33	0.26
	(-2.91)	(-2.87)	(-2.46)	(3.51)	(3.51)	(1.84)
L#INSTN[1]	0.41	0.41	0.39	-0.51	-0.51	0.70
	(6.21)	(6.25)	(5.11)	(-0.87)	(-0.88)	(0.38)
Regression Summary Statistics						
# Observations	1653	1653	1653	1653	1653	1653
Adjusted R^2	0.1010	0.1005	0.0523	0.1442	0.1444	0.1035
Regression F	21.61	24.08	11.34	31.93	35.86	23.72
p-value for F	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001

[1]The lagged number of institutions holding the firm is divided by 100 for this estimation, so the coefficient can be reported in two decimal places.