Industry Information Diffusion and the Lead-lag Effect in Stock Returns

Kewei Hou
Department of Finance, Fisher College of Business, The Ohio State University

I argue that the slow diffusion of industry information is a leading cause of the lead-lag effect in stock returns. I find that the lead-lag effect between big firms and small firms is predominantly an intra-industry phenomenon. Moreover, this effect is driven by sluggish adjustment to negative information, and is robust to alternative determinants of the lead-lag effect. Small, less competitive and neglected industries experience a more pronounced lead-lag effect. The lead-lag effect is related to the post-announcement drift of small firms following the earnings releases of big firms within the industry. (*JEL* G12, G14)

1. Introduction

One of the paramount concerns of financial economics is understanding how firms transmit information to markets, and how markets impound this information into stock prices. Traditional asset-pricing theories assume that in a complete and frictionless market, information diffusion takes place instantaneously. Yet there is ample empirical evidence to suggest that investors face sizeable frictions, and that information can and sometimes does transmit slowly in the market place.

In particular, there is considerable evidence pointing to a lead-lag effect in the equity markets, in which some firms' stock prices show a delayed reaction to price innovations of other firms. In addition, this effect is asymmetric. For example, returns of small firms are correlated with past returns of big firms, but not vice versa [Lo and MacKinlay (1990)].

The article is based on my dissertation at the Graduate School of Business, University of Chicago. I am grateful to the members of my dissertation committee, Nicholas Barberis, George Constantinides, Eugene Fama, Richard Leftwich, and especially Tobias Moskowitz (the chair), for their guidance and encouragement. I also thank Nicole Boyson, Douglas Diamond, Milton Harris, Campbell Harvey (the editor), Owen Lamont, Tim Loughran, Lubos Pastor, David Robinson, Rene Stulz, Alvaro Taboada, an anonymous referee, and seminar participants at the University of Chicago, the University of Texas at Austin, Emory University, MIT, Yale University, the University of North Carolina, Vanderbilt University, the University of Southern California, Cornell University, and the Ohio State University for many helpful comments and suggestions as well as Sandra Sizer for editorial assistance. Research support from the Dice Center for Research in Financial Economics, the Dean's Summer Research Fellowship at the Ohio State University, and the Sanford J. Grossman Fellowship in honor of Arnold Zellner is gratefully acknowledged. Any remaining errors are my responsibility. Address correspondence to Kewei Hou, Department of Finance, Fisher College of Business, Ohio State University, 2100 Neil Avenue, Columbus, OH 43210, or e-mail: hou.28@osu.edu.

doi:10.1093/rfs/hhm003 Advance Access publication January 22, 2007

Although the lead-lag effect can be the result of nonsynchronous trading or time-varying expected returns, several authors [Lo and MacKinlay (1990), Mech (1993), McQueen et al. (1996) and Chordia and Swaminathan (2000)] show that these two explanations can only explain a small portion of the observed lead-lag patterns.

In this article, I primarily focus on a third group of explanations, which argues that the lead-lag effect arises because certain firms react more sluggishly to common information than do others [Lo and MacKinlay (1990), Brennan et al. (1993) and Badrinath et al. (1995)]. This slow diffusion of information can result from many sources, including incomplete markets and limited stock market participation, asymmetric information, noise traders, limited investor attention, transaction costs, short sale constraints, legal restrictions faced by institutional investors, and other types of market frictions and institutional constraints.

To explore the information diffusion hypothesis, I study the lead-lag effect by conditioning on industry membership, because slow diffusion of common information should be more prevalent among firms from the same industry. My results show that there is a strong intra-industry lead-lag effect. Within the same industry, big firms lead small firms. This effect is much more important than the effect across industries. In addition, once I control for the intra-industry effect, the market-wide lead-lag effect becomes insignificant. My intra-industry lead-lag effect findings are robust to alternative empirical specifications and are confirmed for both halves of the sample period.

If market frictions and institutional forces are responsible for the delayed response of stock prices to new information, then because the impact of market imperfections is usually more pronounced when bad news arrives, there might be an asymmetry in the price adjustment process. My results confirm this hypothesis. I find that the intra-industry lead-lag effect is largely driven by slow diffusion of bad news. The ability of big firms' lagged returns to predict small firms' current returns is much greater when big firms' returns are negative.

Previous research [Brennan et al. (1993), Badrinath et al. (1995) and Chordia and Swaminathan (2000)] find that, in addition to size, analyst coverage, institutional ownership, and trading volume are also important determinants of the lead-lag effect. I re-examine these prior results after I condition on industries. My findings confirm that these variables play important roles in determining the lead-lag patterns within industries that are independent of the role played by size.

Next, I examine whether the lead-lag effect is related to the market share of a firm, since new information is usually incorporated into the stock prices of industry leaders before it spreads to other firms within the industry. I find that firms with high market share (industry leaders) lead firms with low market share (industry followers). In addition, when I use the dispersion in analysts' earnings forecasts to measure the uncertainty in the information environment surrounding a firm, I find that firms with less information uncertainty lead firms with greater information uncertainty.

As a robustness check, I compute a delay measure for each individual firm to capture the average price delay with which a firm's stock price responds to common shocks. My results show that size, analyst coverage, institutional ownership, trading volume, market share, and analyst dispersion are all significantly correlated with price delay, and that the directions of the correlations are consistent with the information-based explanation.

I find that the significance of the lead-lag effect varies across industries. The effect is more pronounced in industries that are smaller and less competitive, with lower levels of analyst coverage, institutional ownership and trading volume, and higher levels of analyst dispersion. These are the industries in which I expect, *a priori*, information to diffuse more slowly.

When I examine news in earnings announcements, I find that the lead-lag effect is related to the delayed response of small firms to past earnings surprises on big firms from the same industry. This finding is robust to alternative measures of earnings surprise as well as different empirical techniques.

The rest of the article is organized as follows. Section 2 introduces the data. In Section 3, I show that the lead-lag effect contains a persistent intra-industry component that drives much of the effect. I also show that this effect is largely driven by stock prices' sluggish adjustment to negative information. Section 4 explores additional determinants of the intra-industry lead-lag effect, and in Section 5 I examine cross-industry differences in the lead-lag effect. Section 6 relates the lead-lag effect to stock prices' sluggish adjustment to earnings news. Section 7 concludes.

2. Data

My sample comprises all publicly traded securities with sharecodes 10 or 11 (e.g., excluding ADRs, closed-end funds, and REITs) on the Center for Research in Security Prices (CRSP) NYSE/Amex/Nasdaq data files for the period beginning in July 1963 and ending in December 2001.

I use weekly, as opposed to monthly or daily returns, to estimate the lead-lag effect. Compared to monthly returns, weekly returns offer improved estimation accuracy. Although daily or even intra-daily returns can improve precision further, they may also introduce confounding microstructure influences (such as bid-ask bounce and nonsynchronous trading). Thus, weekly returns seem to be the natural compromise. I calculate weekly returns by compounding daily returns between adjacent Wednesdays as in Hou and Moskowitz (2005).

I then match the weekly return series with balance sheet and income statement data from Compustat, institutional ownership data from Standard & Poor's (S&P), as well as analyst coverage data from Institutional Brokers Estimate System (IBES). The data on institutional ownership are available from 1981 and on, and the data on analyst coverage are available from 1976 and on. They are generally biased towards larger firms.

Book equity is stockholder's equity (Compustat# 216), plus balance sheet deferred tax and investment tax credit (35, if available), minus the book value of preferred stock [liquidating value (10) if available, or else redemption value (56) if available, or else carrying value (130)]. Sales is net sales (12). I match book equity and sales for fiscal year ending in year $t-1$ with CRSP weekly returns from July of year t to June of year $t+1$. In addition, size is CRSP market capitalization (number of shares outstanding times share price) at the end of June of year t. The book-to-market ratio is book equity divided by CRSP market capitalization at the end of December of year $t-1$. I define turnover as the average ratio of the number of shares traded in a week to the number of shares outstanding at the end of the week, averaged from July of year $t-1$ to June of year t. Institutional ownership is measured at the end of year $t-1$. Analyst coverage is the average number of investment analysts following a firm from July of year $t-1$ to June of year t.

3. Intra-industry Lead-lag Effect

If the lead-lag effect is driven by slow diffusion of common information, it should largely be an intra-industry phenomenon. Firms within an industry compete in the same product market and their operating decisions reflect strategic interactions between them. They move closely with each other in product and technology innovations, and react similarly to shifts in supply and demand conditions, as well as changes in regulatory environment. In addition, as the industry experiences expansions and contractions, these firms' growth opportunities and investing and financing decisions are highly correlated. These commonalities suggest that news generated on a given firm is likely to have significant value implications for firms within, rather than outside, its own industry. Consequently, if the lead-lag effect is due to information transmitting slowly across firms, then the clustering of information at the industry level should give rise to a strong lead-lag effect among firms from the same industry.

To study the lead-lag effect within industries, at the end of June of each year from 1963 to 2001 I assign all firms in my sample to one of 12 industries based on their four-digit SIC code, following the industry definitions obtained from Ken French's website. The 12 industries are: (1) consumer nondurables; (2) consumer durables; (3) manufacturing;

(4) oil, gas, and coal extraction and products; (5) chemicals and allied products; (6) business equipment; (7) telephone and television transmission; (8) utilities; (9) wholesale, retail, and some services; (10) healthcare, medical equipment, and drugs; (11) finance; and (12) others. Assigning firms into 12 industries represents a compromise between having a reasonable number of distinct industries and having enough firms within each industry so that sorting within industries will not produce portfolios that are too thin. My results remain largely unchanged when I use alternative industry classification procedures.

I sort firms in each industry into three size portfolios (bottom 30%, middle 40%, and top 30%) according to their end-of-June market capitalization. I then calculate equal-weighted weekly returns for each portfolio from July of year t to June of year $t + 1$.

Table 1 reports summary statistics for the portfolios of the smallest 30% firms (Portfolio 1) and the largest 30% firms (Portfolio 3) from each industry. The average return of small firms is always higher than that of big firms. For instance, the average return of the smallest 30% firms from Industry 3 (Manufacturing) is 41 basis points per week, but it is only 24 basis points per week for the largest 30% firms from Industry 3. The first-order autocorrelation decreases with size—$\rho_1(1, 1)$ is bigger than $\rho_1(3, 3)$ for each of the 12 industries. Higher order autocorrelations also decline with size and decay over time.

3.1 Cross-autocorrelations and VAR test of the lead-lag effect

Table 1 also reports, for each industry, the first- through fourth-order cross-autocorrelations between lagged returns on big firms and current returns on small firms ($\rho_m(1, 3)$, $m = 1$ to 4), and between lagged returns on small firms and current returns on big firms [$\rho_m(3, 1)$]. These correlations show that the lead-lag effect has a significant intra-industry component. The cross-autocorrelations between lagged returns on big firms and current returns on small firms within the same industry are always greater than those between lagged returns on small firms and current returns on big firms.

The asymmetric cross-autocorrelations are consistent with the information-based hypothesis, that the stock prices of small firms react more sluggishly to new information than do the stock prices of big firms. On the other hand, the cross-autocorrelation patterns could also be consistent with an alternative hypothesis based on time-varying expected returns. That is, the cross-autocorrelations between big firms and small firms are the manifestation of the high autocorrelations of small firms combined with the high contemporaneous correlation between big and small firms [Conrad and Kaul (1988, 1989), Boudoukh et al. (1994) and Hameed (1997)]. Under this explanation, the lagged returns of big firms are noisy

Table 1
Summary statistics for intra-industry size portfolios

Industry	Size portfolio (1 = Small, 3 = Big)	N	Mean return	Std. dev. return	Mean size	Median size	Autocorrelations ρ_0 (j,1)	ρ_0 (j,3)	ρ_1 (j,1)	ρ_1 (j,3)	ρ_2 (j,1)	ρ_2 (j,3)	ρ_3 (j,1)	ρ_3 (j,3)	ρ_4 (j,1)	ρ_4 (j,3)
1	1	102	0.34	2.25	0.009	0.008	1.00	0.62	0.35	0.33	0.20	0.17	0.16	0.13	0.10	0.10
	3	102	0.27	1.96	2.122	0.657	0.62	1.00	0.10	0.16	0.06	0.09	0.01	0.07	−0.01	0.01
2	1	45	0.36	2.90	0.010	0.009	1.00	0.61	0.27	0.32	0.17	0.16	0.11	0.11	0.10	0.08
	3	45	0.26	2.44	3.209	0.519	0.61	1.00	0.09	0.19	0.07	0.06	0.02	0.06	0.03	0.03
3	1	193	0.41	2.29	0.011	0.010	1.00	0.71	0.38	0.34	0.22	0.18	0.16	0.15	0.11	0.11
	3	193	0.24	2.30	1.392	0.562	0.71	1.00	0.10	0.18	0.04	0.05	0.01	0.06	−0.00	−0.00
4	1	67	0.56	3.13	0.013	0.012	1.00	0.58	0.26	0.29	0.14	0.12	0.07	0.06	0.06	0.04
	3	67	0.22	2.70	3.452	1.095	0.58	1.00	0.05	0.11	−0.00	−0.00	−0.00	−0.00	0.00	−0.00
5	1	34	0.39	2.71	0.014	0.012	1.00	0.51	0.26	0.25	0.13	0.09	0.11	0.09	0.05	0.05
	3	34	0.25	2.20	3.712	1.340	0.51	1.00	0.08	0.10	0.01	−0.00	0.03	0.03	−0.00	−0.01
6	1	179	0.58	3.13	0.008	0.008	1.00	0.67	0.40	0.38	0.23	0.15	0.16	0.14	0.09	0.11
	3	179	0.28	3.70	1.755	0.345	0.67	1.00	0.10	0.12	0.06	0.03	0.04	0.08	−0.05	−0.02
7	1	25	0.54	3.92	0.026	0.022	1.00	0.49	0.19	0.23	0.11	0.12	0.08	0.13	0.04	0.02
	3	25	0.26	2.63	5.879	1.338	0.49	1.00	0.05	0.07	−0.00	0.01	0.02	0.05	0.00	−0.03
8	1	47	0.28	1.22	0.063	0.057	1.00	0.55	0.26	0.28	0.13	0.13	0.08	0.10	0.03	0.05
	3	47	0.19	1.76	2.323	1.831	0.55	1.00	0.05	0.12	0.00	0.02	−0.02	0.03	−0.02	0.01
9	1	138	0.39	2.29	0.008	0.008	1.00	0.67	0.41	0.36	0.26	0.18	0.21	0.15	0.15	0.11
	3	138	0.28	2.48	1.202	0.353	0.67	1.00	0.13	0.19	0.06	0.07	0.05	0.06	0.02	0.04
10	1	82	0.54	3.03	0.019	0.017	1.00	0.59	0.34	0.29	0.19	0.11	0.14	0.09	0.11	0.11
	3	82	0.30	2.75	2.108	0.682	0.59	1.00	0.03	0.08	0.03	−0.01	−0.01	0.03	−0.02	−0.01
11	1	275	0.37	2.02	0.014	0.014	1.00	0.67	0.37	0.32	0.22	0.17	0.18	0.15	0.11	0.10
	3	275	0.25	1.96	1.305	0.421	0.67	1.00	0.12	0.19	0.03	0.04	0.02	0.04	−0.00	0.01
12	1	165	0.51	2.50	0.009	0.008	1.00	0.68	0.37	0.36	0.21	0.17	0.16	0.14	0.10	0.12
	3	165	0.25	2.54	0.808	0.342	0.68	1.00	0.10	0.17	0.04	0.05	0.02	0.04	−0.02	−0.00

Table 1 reports summary statistics for size-sorted portfolios for each industry from July 1963 to December 2001. At the end of June of each year t, I assign firms on CRSP NYSE/Amex/Nasdaq tapes with sharecodes 10 or 11 into one of 12 industry portfolios based on their four-digit SIC code. I obtain the industry definitions from Ken French's website. I then sort firms within each industry portfolio into three size portfolios (bottom 30%, middle 40%, and top 30%) according to their end-of-June market capitalization. I compute equal-weighted weekly returns for each portfolio from July of year t to June of year $t+1$. Portfolio 1 refers to the portfolio of the smallest 30% firms, and Portfolio 3 refers to the portfolio of the largest 30% firms. N is the average number of firms in each portfolio. I calculate mean and median size (in billions of dollars) first across firms within each portfolio and then average across time. $\rho_m(j,k), m = 0$ to 4, refers to the m^{th} order correlation coefficient between returns on the j^{th} size-ranked portfolio ($j = 1$ or 3) and returns on the k^{th} size-ranked portfolio ($k = 1$ or 3). For example, $\rho_1(1,3)$ represents the correlation between week t return on the small size portfolio and week $t-1$ return on the big size portfolio, and $\rho_1(3,1)$ represents the correlation between week t return on the big size portfolio and week $t-1$ return on the small size portfolio. Asymptotic standard errors for the correlation coefficients are equal to 0.0223 under the i.i.d. null.

proxies for the lagged returns of small firms and, once I control for lagged small-firm returns, the lead-lag effect will disappear.

To address this alternative hypothesis and test the lead-lag effect between big and small firms within the same industry formally, I estimate the following vector autoregressions (VARs) jointly across all 12 industries:

$$R_{i,1}(t) = a_{i,0} + \sum_{k=1}^{K} a_k R_{i,1}(t-k) + \sum_{k=1}^{K} b_k R_{i,3}(t-k) + e_{i,1}(t), \qquad (1)$$

$$R_{i,3}(t) = c_{i,0} + \sum_{k=1}^{K} c_k R_{i,1}(t-k) + \sum_{k=1}^{K} d_k R_{i,3}(t-k) + e_{i,3}(t). \qquad (2)$$

In Equations (1) and (2), $R_{i,1}(t)$ and $R_{i,3}(t)$ are the week t returns on the portfolios of the smallest 30% and the largest 30% firms from industry i, respectively. Similar regression models are used in Brennan et al. (1993) and Chordia and Swaminathan (2000) to study the lead-lag relations between stocks with different levels of analyst coverage and trading volume.

To better capture the average effect across all industries, I restrict the autoregressive and cross-autoregressive coefficients to be identical across industries in the joint estimation. However, my results are robust to estimating the VARs for each industry separately, without imposing the coefficient restrictions and then testing for the lead-lag effect jointly across industries.

I estimate Equations (1) and (2) with one ($K = 1$) and four ($K = 4$) lags. The advantage of using one lag is that it is easy to interpret, but it does not allow for time-series dependency in returns beyond one week. Using four lags, although possibly adding noise to the estimation procedure, eliminates this problem to a large extent, as it seems reasonable to assume that small firms will react to information about big firms within a month's time. I have introduced higher lags to Equations (1) and (2) and found nearly identical results.

If the lead-lag effect between big firms and small firms is the result of small firms' own autocorrelations coupled with high contemporaneous correlation between big and small firms, then once I control for the lagged returns of small firms, the cross-autoregressive coefficients in the VARs should be indistinguishable from zero. On the other hand, if the lead-lag effect is driven by the stock prices of big firms responding more rapidly to common information than those of small firms, then lagged returns on big firms should continue to predict current returns on small firms. In addition, this predictive power should be greater than that of lagged returns on small firms for current returns on big firms. In the context of the VARs in Equations (1) and (2), the information-based hypothesis suggests that the sum of the coefficients on $R_{i,3}(t-k)$ in Equation (1) should be significantly different from zero and greater than the sum of the coefficients on $R_{i,1}(t-k)$ in Equation (2), i.e., $\sum_{k=1}^{K} b_k > \sum_{k=1}^{K} c_k$.

Table 2, Panel A, presents the VAR estimation results. The sum of b_k is positive and statistically significant at the 1% level for both the four-lag (0.26, F-statistic = 302.30) and one-lag (0.18, t-statistic = 24.02) regressions. Thus, lagged returns of big firms do contain information about future returns of small firms beyond that contained in lagged returns of small firms.

Moreover, the economic magnitude of the effect is large. A 1% decrease in last week's return of big firms leads to an 18 basis points decrease in the return of small firms this week, all else being equal. In Table 1, the average standard deviation of the return on big firms is 2.45%. Therefore, a one standard deviation decrease in the return of big firms, on average, leads to a 44 basis points decrease in the return of small firms within the same industry.

Meanwhile, the sum of c_k is not different from zero in either the one-lag or four-lag VARs, and the F-statistic easily rejects the cross-equation restriction of $\sum_{k=1}^{K} b_k = \sum_{k=1}^{K} c_k$ at the 1% level. These results indicate a lead-lag relation between big firms and small firms within the same industry that is both statistically and economically significant.

3.2 Robustness checks

My results are robust to alternative VAR specifications, potential microstructure biases, subsamples and subperiods, and value weighting in returns. To conserve space, in Panel B of Table 2 I report only a subset of these robustness findings. The first set of numbers are obtained by estimating the VARs for each industry separately and then averaging the coefficients across industries. The average coefficients are almost identical to the coefficients reported in Panel A, confirming that the specification in Equations (1) and (2) is useful for capturing the average effect across industries. The average sum of b_k is significant at the 1% level for both the one-lag and four-lag regressions. On the other hand, the average sum of c_k is not different from zero, and the cross-equation restriction that the average sum of b_k equals the average sum of c_k is rejected at the 1% level.

In the next specification, I skip a week in the VAR estimation (that is, the explanatory variables are now weekly returns from week $t-5$ to week $t-2$) to make sure that my results are not driven by microstructure effects such as nonsynchronous trading or bid-ask bounce. The estimated coefficients show a sizable reduction, which is not surprising since most of the significance in the VARs occurs at the 1-week lag. Nevertheless, the lead-lag effect remains robust. The sum of b_k is still highly statistically significant. Further, the hypothesis that it is equal to the sum of c_k is again

Table 2
Test of intra-industry lead-lag effect

Panel A: Intra-industry lead-lag effect

	Four-lag regressions			One-lag regressions		
LHS	$R_{i,1}(t-1:t-4)$	$R_{i,3}(t-1:t-4)$	F_1	$R_{i,1}(t-1)$	$R_{i,3}(t-1)$	F_1
$R_{i,1}(t)$	0.30	0.26	204.37***	0.21	0.18	309.77***
	*565.55****	*302.30****		*28.21****	*24.02****	
$R_{i,3}(t)$	−0.02	0.19		−0.00	0.14	
	2.14	*150.23****		*−0.42*	*17.80****	

Panel B: Intra-industry lead-lag effect, robustness checks

	Four-lag regressions			One-lag regressions		
LHS	$R_{i,1}(t-1:t-4)$	$R_{i,3}(t-1:t-4)$	F_1	$R_{i,1}(t-1)$	$R_{i,3}(t-1)$	F_1
		Unconstrained estimation				
$R_{i,1}(t)$	0.29	0.26	207.50***	0.21	0.18	293.13***
	*507.49****	*289.86****		*764.76****	*531.14****	
$R_{i,3}(t)$	−0.03	0.19		−0.01	0.14	
	2.44	*143.18****		*0.95*	*312.11****	
		Skipping 1 week				
$R_{i,1}(t)$	0.19	0.16	79.53***	0.16	0.07	15.30***
	*192.09****	*105.16****		*19.50****	*7.03****	
$R_{i,3}(t)$	−0.02	0.09		0.02	0.04	
	2.06	*29.79****		*2.24***	*4.16****	
		196307–198206				
$R_{i,1}(t)$	0.25	0.36	163.34***	0.17	0.24	214.97***
	*162.33****	*202.48****		*15.02****	*18.31****	
$R_{i,3}(t)$	−0.03	0.29		0.00	0.17	
	*2.87**	*154.81****		*0.27*	*14.68****	
		198207–200112				
$R_{i,1}(t)$	0.34	0.20	77.90***	0.25	0.15	139.96***
	*412.44****	*121.40****		*24.35****	*16.88****	
$R_{i,3}(t)$	−0.03	0.13		−0.03	0.12	
	*2.95**	*32.74****		*−1.84**	*11.07****	

Panel C: Intra-versus inter-industry lead-lag effects

	Four-lag regressions				One-lag regressions			
LHS	$R_{i,1}$ $(t-1:t-4)$	$R_{i,3}$ $(t-1:t-4)$	$R_{j\neq i,3}$ $(t-1:t-4)$	$F_1(F_2)$	$R_{i,1}$ $(t-1)$	$R_{i,3}$ $(t-1)$	$R_{j\neq i,3}$ $(t-1)$	$F_1(F_2)$
$R_{i,1}(t)$	0.30	0.28	0.02	128.18***	0.21	0.17	0.02	158.15***
	*513.28****	*159.38****	*0.43*	(41.12***)	*26.62****	*15.21****	*1.91**	(40.21***)
$R_{i,3}(t)$	−0.02	0.19			−0.00	0.14		
	2.14	*150.23****			*−0.42*	*17.80****		

Panel D: Market-wide lead-lag effect, controlling for intra-industry effect

	Four-lag regressions			One-lag regressions		
LHS	$R_1(t-1:t-4)$	$R_3(t-1:t-4)$	F_1	$R_1(-1)$	$R_3(t-1)$	F_1
R_1(t)	0.15	−0.05	0.01	0.09	−0.03	0.21
	*6.23***	*0.68*		*2.53****	*−0.81*	
R_3(t)	−0.04	−0.02		−0.05	0.03	
	0.35	*0.09*		*−1.28*	*0.89*	

Table 2
(Continued)

Panel E: Intra-industry lead-lag effect, good news versus bad news

LHS	$R_{i,1}(t-1)D_{i,1}$	$R_{i,1}(t-1)$	$R_{i,3}(t-1)D_{i,3}$	$R_{i,3}(t-1)$
$R_{i,1}(t)$	0.14	0.10	−0.16	0.28
	6.60***	*6.03****	*−6.86****	*17.83****
$R_{i,3}(t)$	0.01	−0.01	−0.01	0.16
	0.41	*−0.59*	*−0.57*	*9.88****

In Panel A, I estimate the following four-lag ($K = 4$) and one-lag ($K = 1$) vector autoregressions (VARs) jointly across all industries, using weekly returns on the size portfolios formed within each industry, from July 1963 to December 2001:

$$R_{i,1}(t) = a_{i,0} + \sum_{k=1}^{K} a_k R_{i,1}(t-k) + \sum_{k=1}^{K} b_k R_{i,3}(t-k) + e_{i,1}(t), \tag{T.2.1}$$

$$R_{i,3}(t) = c_{i,0} + \sum_{k=1}^{K} c_k R_{i,1}(t-k) + \sum_{k=1}^{K} d_k R_{i,3}(t-k) + e_{i,3}(t). \tag{T.2.2}$$

In Equations (T.2.1) and (T.2.2), $R_{i,1}(t)$ is the week t return on the portfolio of the smallest 30% firms from industry $i. R_{i,3}(t)$ is the week t return on the portfolio of the biggest 30% firms from industry i. I restrict the autoregressive and cross-autoregressive coefficients to be identical across industries in the joint estimation. $R_{i,1}(t-1:t-k), k = 1$ or 4, reports $\sum_{k=1}^{K} a_k$ from Equation (T.2.1) or $\sum_{k=1}^{K} c_k$ from Equation (T.2.2), depending on the left-hand side variable. Similarly, $R_{i,3}(t-1:t-k), k = 1$ or 4, reports $\sum_{k=1}^{K} b_k$ from Equation (T.2.1) or $\sum_{k=1}^{K} d_k$ from Equation (T2.2). *Italics* indicate the F-Statistics (t-statistics) for the hypothesis that the sum of the coefficients equals zero in the four-lag (one-lag) regressions. F_1 refers to the F-statistic for the cross-equation hypothesis that $R_{i,3}(t-1:t-k)$ from Equation (T.2.1) equals $R_{i,1}(t-1:t-k)$ from Equation (T.2.2), i.e., $\sum_{k=1}^{K} b_k = \sum_{k=1}^{K} c_k$. In Panel B, as robustness checks, I estimate the four- and one-lag VARs for each industry separately without the cross-industry coefficient restrictions. I report the average sums of coefficients across industries and the corresponding F-statistics that they equal zero. The rest of the panel re-estimates the specification in Panel A by skipping a week and for two subperiods (196307–198206 and 198207–200112). In Panel C, for industry i, I add lagged returns on an equal-weighted portfolio of big firms from outside industry i [$R_{j\neq i,3}(t-k), k = 1$ to 4] to the right-hand side of Equation (T.2.1):

$$R_{i,1}(t) = a_{i,0} + \sum_{k=1}^{K} a_k R_{i,1}(t-k) + \sum_{k=1}^{K} b_k R_{i,3}(t-k) + \sum_{k=1}^{K} f_k R_{j\neq i,3}(t-k) + e_{i,1}(t), \tag{T.2.3}$$

$$R_{i,3}(t) = c_{i,0} + \sum_{k=1}^{K} c_k R_{i,1}(t-k) + \sum_{k=1}^{K} d_k R_{i,3}(t-k) + e_{i,3}(t). \tag{T.2.4}$$

$R_{j\neq i,3}(t-1:t-k)$ reports the sum of the corresponding regression coefficients from Equation (T.2.3). F_2 refers to the F-statistic for the hypothesis $\sum_{k=1}^{K} b_k = \sum_{k=1}^{K} f_k$. Panel D estimates the market-wide lead-lag effect using returns that have been pre-adjusted for the intra-industry effect. To account for the lead-lag effect within industries, I first regress weekly stock returns of individual firms on the past 4 weeks' returns of the large and small size portfolios for the industry to which each firm belongs. I then use the residuals from the firm-level lead-lag regressions to estimate four- and one-lag VARs on size portfolios formed across firms irrespective of their industry membership. In Panel E, I add dummy variables for positive and negative lagged returns to the one-lag VARs.

$$R_{i,1}(t) = a_{i,0} + a_1 R_{i,1}(t-1)D_{i,1} + a_2 R_{i,1}(t-1) + b_1 R_{i,3}(t-1)D_{i,3} + b_2 R_{i,3}(t-1) + e_{i,1}(t), \tag{T.2.5}$$

$$R_{i,3}(t) = c_{i,0} + c_1 R_{i,1}(t-1)D_{i,1} + c_2 R_{i,1}(t-1) + d_1 R_{i,3}(t-1)D_{i,3} + d_2 R_{i,3}(t-1) + e_{i,3}(t). \tag{T.2.6}$$

In Equation (T.2.5) and (T.2.6), $D_{i,1}$ equals one if $R_{i,1}(t-1)>0$, and zero otherwise. Similarly, $D_{i,3}$ equals one if $R_{i,3}(t-1)>0$, and zero otherwise. ***, **, and * denote significance at the 1, 5, and 10% levels, respectively.

strongly rejected at the 1% level, suggesting that microstructure issues have little impact on my results.

I also split the entire sample period into two subperiods (July 1963 to June 1982 and July 1982 to December 2001). The lead-lag effect is highly significant in both subperiods, although it is weaker for the second half of the sample. This result is consistent with the information-delay interpretation of the lead-lag effect, because the second half of the sample coincides with a period in which relaxation of various market frictions and investment restrictions, increased disclosure, and improvement in information communication and market mechanism help stock prices become informationally more efficient.

In unreported tables, I replicate my findings using value-weighted returns instead of equal-weighted returns on the size portfolios. Doing so biases my results towards larger firms. It is also likely to further reduce the impact of microstructure effects, which are more prevalent among smaller firms. In addition, to ensure that my findings on the intra-industry lead-lag effect are not entirely driven by the small firms on Amex and Nasdaq, I confirm my results using only NYSE firms. I also find that the market-wide lead-lag effect becomes insignificant when I use either value-weighted returns or only NYSE firms. The equal-weighted market-wide effect for the entire sample (NYSE/Amex/Nasdaq), although statistically significant, is of far smaller magnitude when compared to the intra-industry effect in Panels A and B. These findings further emphasize the importance of industry in understanding the source of the lead-lag effect.

3.3 Intra-industry versus inter-industry lead-lag effects

If the slow diffusion of industry-specific information is the primary source of the lead-lag effect between big firms and small firms, then not only should there be a significant lead-lag effect within industries, this effect should also be stronger than the lead-lag effect across industries.

To investigate this hypothesis, for every industry i I add the lagged returns of big firms outside the industry to Equation (1) and estimate the following regression system jointly across industries:

$$\begin{aligned} R_{i,1}(t) &= a_{i,0} + \sum_{k=1}^{K} a_k R_{i,1}(t-k) + \sum_{k=1}^{K} b_k R_{i,3}(t-k) \\ &\quad + \sum_{k=1}^{K} f_k R_{j\neq i,3}(t-k) + e_{i,1}(t), \end{aligned} \tag{3}$$

$$R_{i,3}(t) = c_{i,0} + \sum_{k=1}^{K} c_k R_{i,1}(t-k) + \sum_{k=1}^{K} d_k R_{i,3}(t-k) + e_{i,3}(t). \tag{4}$$

In Equation (3), $R_{j\neq i,3}(t-k)$ is the week $t-k$ return on the portfolio of big firms outside industry i. To ensure that those big firms are strictly larger than the small firms from industry i, I use the size breakpoints for industry i to sort firms from the other 11 industries into three size portfolios. A stronger lead-lag effect within industries would imply that $\sum_{k=1}^{K} b_k > \sum_{k=1}^{K} f_k$ in Equation (3).

Table 2, Panel C, reports the regression results. The lead-lag effect between big firms and small firms within the same industry remains strong and highly significant even after I control for the lagged returns of big firms from outside the industry. The sum of b_k remains virtually unchanged for both the four-lag and one-lag regressions, and the F-statistic for the cross-equation restriction ($\sum_{k=1}^{K} b_k = \sum_{k=1}^{K} c_k$) again easily rejects the null at the 1% level.

More important, it is clear that big firms within an industry dominate those from outside the industry in predicting returns of small firms. Examining the four-lag regressions, for example, the sum of b_k is 0.28 and significant at the 1% level, but the sum of f_k is only 0.02 and statistically insignificant, and the F-statistic rejects the hypothesis of $\sum_{k=1}^{K} b_k = \sum_{k=1}^{K} f_k$ at the 1% level. Therefore, once I control for the lagged returns of big firms within an industry, the lagged returns of big firms outside the industry have no predictive power for current returns of small firms. This result is consistent with the hypothesis that the lead-lag effect is driven by intra-industry news dissemination.

3.4 Intra-industry versus market-wide lead-lag effects

Next, I control for the lead-lag effect within industries, and then test whether there is any remaining effect at the market level. Doing so allows me to directly assess the contribution of the intra-industry effect to the overall lead-lag effect.[1]

To account for the lead-lag effect within industries, I regress the weekly stock returns of individual firms on the past 4 weeks' returns of the big and small size portfolios from the industry to which each firm belongs. Then I use the residuals from the firm-level lead-lag regressions to estimate a market-wide lead-lag effect on size portfolios formed across all firms, irrespective of their industry membership. Any predictability between the residual returns of big and small firms would indicate a market-driven lead-lag effect above and beyond the industry effect.

The results, which I report in Panel D of Table 2, suggest that there is no residual market-wide lead-lag effect once I control for the industry effect. Both the sum of b_k and the sum of c_k are statistically indistinguishable from zero, and the F-statistics cannot reject the hypothesis that they are

[1] I wish to thank the referee for suggesting this test.

equal to each other.[2] Therefore, the overall lead-lag effect appears to be largely driven by the effect within industries.

I also estimate a trivariate VAR system across industries by including market returns as an additional state variable. I find that after I control for the lagged returns on big firms from the same industry, the lagged market returns cannot predict the returns on small firms. On the other hand, the lagged returns on big firms can significantly predict the market returns. For brevity, I do not tabulate these results.

3.5 Does bad news travel slowly?

Since certain market imperfections become more pronounced when bad news arrives (e.g., the short sale constraints can delay the incorporation of negative information into stock prices [Diamond and Verrecchia (1987)]), I explore whether there is an asymmetry in the response of small firms to past returns on big firms. To do so, I include dummy variables in the one-lag VARs to allow the predictive power of a lagged return to differ depending on its sign:

$$\begin{aligned} R_{i,1}(t) &= a_{i,0} + a_1 R_{i,1}(t-1) D_{i,1} + a_2 R_{i,1}(t-1) + b_1 R_{i,3}(t-1) D_{i,3} \\ &\quad + b_2 R_{i,3}(t-1) + e_{i,1}(t), \qquad (5) \\ R_{i,3}(t) &= c_{i,0} + c_1 R_{i,1}(t-1) D_{i,1} + c_2 R_{i,1}(t-1) + d_1 R_{i,3}(t-1) D_{i,3} \\ &\quad + d_2 R_{i,3}(t-1) + e_{i,3}(t). \qquad (6) \end{aligned}$$

In Equations (5) and (6), $D_{i,1}$ or $D_{i,3}$ takes the value of one if last week's return of small firms $[R_{i,1}(t-1)]$ or big firms $[R_{i,3}(t-1)]$ is positive and zero otherwise.

A negative b_1 estimate from Equation (5) would indicate that the predictive power of last week's return on big firms is stronger when the return is negative, and thus support the hypothesis that small firms react more slowly to bad news on big firms. But a zero or positive b_1 estimate would suggest that there is no asymmetry in the speed of adjustment of small firms to past returns on big firms, or that small firms, actually, react more slowly to good news on big firms.

In Panel E of Table 2 I report the regression results. b_1 is -0.16 with a t-statistic of -6.86, and the sum of b_1 and b_2 (0.12) is less than half of b_2 (0.28), indicating that the ability of the lagged returns on big firms to predict current returns on small firms is much greater when big firms experience negative returns in the week before. Therefore, good news is diffused quickly between big firms and small firms within the same

[2] For comparison, when I estimate the market-wide lead-lag effect using raw instead of industry-adjusted returns, the four-lag VARs produce a sum of b_k equal to 0.16, which is significantly different from zero at the 1% level, an insignificant sum of c_k equal to -0.03, and a F-statistic that rejects the cross-equation restriction at the 1% level. The results for the one-lag VARs are similar.

industry, and it is mainly the slow adjustment of small firms' stock prices to bad news on big firms that drives the observed lead-lag effect.

4. Alternative Determinants of Intra-industry Lead-lag Effect

4.1 Analyst coverage, institutional ownership, and trading volume

A number of researchers argue that size affects the speed of stock price adjustment because of its correlations with variables such as analyst coverage [Brennan et al. (1993)], institutional ownership [Badrinath et al. (1995)], and trading volume [Chordia and Swaminathan (2000)]. Since analyst coverage, institutional holdings, and trading volume tend to cluster at the industry level, I examine their roles in determining the lead-lag patterns within industries.

Because analyst coverage, institutional ownership, and trading volume are all positively correlated with size, I control for size when I look for the independent lead-lag effects associated with these variables. Within each industry, I first sort firms into three portfolios (top 30%, middle 40%, and bottom 30%) according to their size. I then subdivide each size portfolio into three portfolios (top 30%, middle 40%, and bottom 30%) according to analyst coverage, institutional ownership, or trading volume.[3] As a result of this two-way sort, I have nine portfolios within each industry.

I then place firms with the same characteristic ranking from each of the three size groups into one portfolio. This procedure gives me, within each industry, three characteristic-ranked portfolios while holding size approximately constant. I use the returns on the highest and lowest characteristic-ranked portfolios ($R_{i,3}$ and $R_{i,1}$, respectively) from each industry to estimate Equations (1) and (2) jointly across industries. The estimation results, which I report in Panels A-C of Table 3, show that there are strong intra-industry lead-lag effects related to analyst coverage, institutional ownership, and trading volume, and that these effects are independent of the size effect. Firms with higher levels of analyst coverage (institutional ownership, or trading volume) lead firms with lower levels of analyst coverage (institutional ownership, or trading volume) from the same industry, even after I control for size.

4.2 Market share

I propose market share as another proxy of the speed of information flow, since there might be a lead-lag relation between industry leaders and other firms within the industry. A new piece of information usually impacts the

[3] In forming these portfolios, I assume that firms that have no IBES analyst coverage or S&P institutional ownership data have zero analyst or institutional following. My results are robust when I exclude these firms. Further, because the reported trading volume on Nasdaq includes interdealer trades but NYSE/Amex does not, I follow other studies such as LaPlante and Muscarella (1997) and Loughran and Schultz (2005) by doubling the volume for NYSE/Amex firms to make that volume comparable to Nasdaq volume. I also repeat the volume-based tests using only NYSE/Amex firms, and find similar results.

Table 3
Alternative determinants of intra-industry lead-lag effect

LHS	Four-lag regressions $R_{i,1}$ $(t-1:t-4)$	Four-lag regressions $R_{i,3}$ $(t-1:t-4)$	Four-lag regressions F_1	One-lag regressions $R_{i,1}$ $(t-1)$	One-lag regressions $R_{i,3}$ $(t-1)$	One-lag regressions F_1
		Panel A: Analyst coverage				
$R_{i,1}(t)$	−0.01	0.25	22.69***	0.03	0.18	32.99***
	0.02	*47.76******		*1.66**	*9.87****	
$R_{i,3}(t)$	0.00	0.24		0.03	0.16	
	0.01	*41.72****		*1.71**	*8.87****	
		Panel B: Institutional ownership				
$R_{i,1}(t)$	0.02	0.25	10.38***	−0.01	0.19	54.24***
	0.42	*72.84****		*−0.58*	*13.43****	
$R_{i,3}(t)$	0.11	0.13		0.04	0.13	
	*11.56****	*16.65****		*2.32***	*8.88****	
		Panel C: Turnover				
$R_{i,1}(t)$	0.08	0.24	24.71***	0.03	0.18	53.90***
	*17.39****	*293.22****		*2.68****	*27.41****	
$R_{i,3}(t)$	0.08	0.22		0.07	0.17	
	*8.22****	*123.02****		*5.34****	*18.08****	
		Panel D: Sales				
$R_{i,1}(t)$	0.15	0.25	36.42***	0.09	0.20	83.79***
	*57.68****	*126.60****		*8.67****	*17.60****	
$R_{i,3}(t)$	0.07	0.25		0.06	0.17	
	*16.19****	*150.91****		*6.69****	*16.27****	
		Panel E: Inverse of analyst dispersion				
$R_{i,1}(t)$	0.09	0.16	12.21***	0.05	0.16	23.00***
	*10.19****	*30.91****		*3.95****	*9.80****	
$R_{i,3}(t)$	0.02	0.13		0.04	0.09	
	0.58	*22.04****		*3.59****	*6.26****	

At the end of June of each year t from 1963 to 2001 (my sample period for analyst coverage starts in 1976 and in 1981 for institutional ownership), I assign firms into one of 12 industry portfolios based on their four-digit SIC code. I obtain the industry definitions from Ken French's website. I first sort firms within each industry portfolio into three size-ranked portfolios (bottom 30%, middle 40%, and top 30%) and within each size portfolios into three characteristic-ranked portfolios (bottom 30%, middle 40%, and top 30%). Then I place firms from the three size portfolios that have the same characteristic ranking into one portfolio. Within each industry, this procedure generates three characteristic-ranked portfolios while holding size fixed. I compute equal-weighted weekly returns for each portfolio from July of year t to June of year $t+1$. Analyst coverage is the average number of analysts following a firm from July of year $t-1$ to June of year t. Institutional ownership is measured in December of year $t-1$. Turnover is the average share turnover per week, which I define as the ratio of the number of shares traded in a week to the number of shares outstanding at the end of the week, averaged from July of year $t-1$ to June of year t. I double the turnover for NYSE/Amex firms so that it is comparable with Nasdaq turnover. Sales is net sales for fiscal year ending in year $t-1$, as reported by Compustat. I define analyst dispersion as the standard deviation of analysts' annual earnings forecast divided by the absolute value of the mean forecast each month averaged over the previous year. I estimate the following four-lag ($K=4$) and one-lag ($K=1$) vector autoregressions (VARs) jointly across all industries:

$$R_{i,1}(t) = a_{i,0} + \sum_{k=1}^{K} a_k R_{i,1}(t-k) + \sum_{k=1}^{K} b_k R_{i,3}(t-k) + e_{i,1}(t), \tag{T.3.1}$$

$$R_{i,3}(t) = c_{i,0} + \sum_{k=1}^{K} c_k R_{i,1}(t-k) + \sum_{k=1}^{K} d_k R_{i,3}(t-k) + e_{i,3}(t). \tag{T.3.2}$$

In Equations (T.3.1) and (T.3.2), $R_{i,1}(t)$ is the week t return on the portfolio of the lowest-ranked 30% firms from industry i. $R_{i,3}(t)$ is the week t return on the portfolio of the highest ranked 30% firms of industry i. I restrict autoregressive and cross-autoregressive coefficients to be identical across industries. $R_{i,1}(t-1:t-k)$, $k=1$ or 4, reports $\sum_{k=1}^{K} a_k$ from Equation (T.3.1) or $\sum_{k=1}^{K} c_k$ from Equation (T.3.2), depending on the right-hand side variable. Similarly, $R_{i,3}(t-1:t-k)$, $k=1$ or 4, reports $\sum_{k=1}^{K} b_k$ in Equation (T.3.1) or $\sum_{k=1}^{K} d_k$ in Equation (T.3.2). *Italics* indicate the F-Statistics (t-statistics) for the hypothesis that the sum of the coefficients equals zero in the four-lag (one-lag) VARs. F_1 refers to the F-statistic for the cross-equation hypothesis that $R_{i,3}(t-1:t-k)$ from Equation (T.3.1) equals $R_{i,1}(t-1:t-k)$ from Equation (T.3.2), i.e., $\sum_{k=1}^{K} b_k = \sum_{k=1}^{K} c_k$. ***, **, and * denote significance at the 1, 5, and 10% levels, respectively.

industry leaders, those firms with high market share, first. However, due to market frictions, this information may not be instantaneously impounded into the prices of other firms in the industry. As a result, a lead-lag relation between industry leaders and followers arises.

To investigate whether there is an independent lead-lag effect related to market share, I again control for size, since the two variables are likely to be highly correlated. Within each industry, I first sort firms into three size portfolios, and then subdivide each size portfolio into three portfolios based on sales. I group firms with the same sales ranking from each of the three size portfolios into one portfolio, resulting in three sales-ranked portfolios holding size fixed.

Panel D of Table 3 presents results of the VAR test using returns on the two extreme sales-ranked portfolios from each industry. These results show that, holding size constant, firms with high market share lead firms with low market share. For both the four-lag and one-lag regressions, the sum of b_k from Equation (1) is greater than the sum of c_k from Equation (2), and the cross-equation restriction is rejected at the 1% significance level.

4.3 Analyst dispersion

I hypothesize that the dispersion of analysts' earnings forecasts (analyst dispersion) also contains information about the lead-lag effect within industries. The analyst dispersion variable is used by past studies [e.g., Imhoff and Lobo (1992), Barron et al. (1998), Barron and Stuerke (1998) and Zhang (2006)] to capture the uncertainty in the information environment surrounding a firm. The higher the analyst dispersion, the noisier the information environment. In addition, many investors, especially institutional investors, tend to avoid stocks with greater information uncertainty. Thus, there could be a slower diffusion of new information into the stock prices of firms with higher levels of analyst dispersion, and the returns of these firms could lag those of their industry peers.

I define analyst dispersion as the standard deviation of analysts' annual earnings forecasts divided by the absolute value of the mean forecast each month, averaged over the previous year.[4] Within each industry, I sort stocks into three portfolios (top 30%, middle 40%, and bottom 30%) based on the inverse of analyst dispersion while holding size constant. I use weekly returns on the two extreme portfolios to conduct the intra-industry lead-lag test.

The results in Panel E of Table 3 demonstrate that holding size constant, firms with lower levels of analyst dispersion lead their industry peers with higher levels of analyst dispersion. Therefore, the evidence supports my

[4] The calculation of analyst dispersion requires at least two valid forecasts to return a nonmissing value. Thus, I exclude firms that are followed by fewer than two analysts from the test. Doing so introduces a bias toward larger and more visible firms.

hypothesis that analyst dispersion is related to the speed of information diffusion.

4.4 Robustness checks using individual stock's price delay

To investigate the various lead-lag determinants in a unified framework and to ensure robustness of the VAR findings, I extend Mech (1993), Chordia and Swaminathan (2000) and Hou and Moskowitz (2005) and compute the average delay with which a firm's stock price responds to market and industry shocks. At the end of June of each year, I regress each individual stock's weekly returns on contemporaneous and the past 4 weeks' returns on the market and industry portfolios over the previous 5 years:

$$r_{j,t} = \alpha_{\mathrm{j}} + \beta_0 R_{M,t} + \delta_0 R_{I,t} + +\varepsilon_{j,t}, \tag{7}$$

$$r_{j,t} = \alpha_{\mathrm{j}} + \beta_0 R_{M,t} + \sum_{\mathrm{n}=1}^{4} \beta_n R_{M,t-n} + \delta_0 R_{I,t} + \sum_{\mathrm{n}=1}^{4} \delta_n R_{I,t-n} + \varepsilon_{j,t}. \tag{8}$$

In Equations (7) and (8), $r_{j,t}$ is the week t return of firm j, and $R_{M,t}$ and $R_{I,t}$ are the returns on the CRSP value-weighted market portfolio and the industry portfolio to which firm j belongs, respectively. My price delay measure is then the fraction of firm j's return variation explained by the lagged market and industry returns. It is equal to one minus the ratio of the R-squared from regression (7) to the R-squared from regression (8).

Then, for each year I estimate a cross-sectional regression of individual firms' price delay on the various lead-lag determinants examined in the previous sections. Because the R-squared-based delay measure is bounded between zero and one, I perform a logistic transformation [*ln(DELAY/(1-DELAY)*] before running the regressions. This econometric treatment has the added benefit of removing the excess skewness and kurtosis of the original price delay measure while preserving its monotonicity. The regressions' explanatory variables are the natural logarithms of size, turnover, one plus analyst coverage, one plus institutional ownership, market share, and analyst dispersion. Due to the data availability on analyst coverage and institutional ownership, the annual regressions start in 1981. However, my results are robust when I exclude those variables from the regressions and extend the sample back to 1963.

Equation (9) below reports the time-series averages of the cross-sectional regression coefficients and their time-series t-statistics:

$$\begin{aligned} logit\ (DELAY) = &\underset{(-12.76)}{-0.26\ \ln(size)} \quad \underset{(-6.52)}{-\ 0.26\ \ln(1+\#Analyst)} \\ &\underset{(-4.48)}{-\ 0.35\ \ln(1+InstitutionalOwnership)} \quad \underset{(-16.24)}{-\ 0.31\ \ln(Turnover)} \\ &\underset{(-8.06)}{-\ 0.11\ \ln(MktShare)} \quad \underset{(3.48)}{+\ 0.04\ \ln(AnalystDispersion)}. \end{aligned} \tag{9}$$

The regressions corroborate the lead-lag findings in Tables 2 and 3 very nicely. Delay loads negatively on size, analyst coverage, institutional ownership, turnover, and market share, and positively on analyst dispersion. These results suggest that firms whose stock prices adjust with a delay to market- and industry-wide news are small, neglected by analysts and institutional investors, have low trading volume and market share, and high analyst dispersion. These results provide additional support to the information-based interpretations, because the most delayed firms are those that Tables 2 and 3 show as lagging their industry peers.

5. Cross-industry Variation in Lead-lag Effect

Here, I investigate the variation of the lead-lag effect across industries, since many of the factors that cause information to transmit slowly also differ substantially at the industry level. I am particularly interested to learn how the significance of the intra-industry lead-lag effect relates to the industry values of the set of variables that affect the speed of price adjustment at the firm level.

I begin by adding an interaction term between the lagged return on big firms $[R_{i,3}(t-1)]$ and the characteristic of industry i (IC_i) to Equation (1) of the one-lag VARs:

$$R_{i,1}(t) = a_{i,0} + a_1 R_{i,1}(t-1) + b_1 R_{i,3}(t-1) + b_2 R_{i,3}(t-1) \times IC_i + e_{i,1}(t), \tag{10}$$

$$R_{i,3}(t) = c_{i,0} + c_1 R_{i,1}(t-1) + d_1 R_{i,3}(t-1) + e_{i,3}(t). \tag{11}$$

This setup allows the predictive power of the lagged big-firm return to vary across industries with the industry characteristic of interest. For example, a positive coefficient estimate for the interaction term (b_2) would suggest that the lead-lag effect is more pronounced in industries for which the value of IC is higher. I estimate Equations (10) and (11) jointly across industries and report the regression coefficients in Table 4.

Panel A of Table 4 uses industry average size as the relevant industry characteristic in Equation (10). The results confirm a positive relation between size and the speed of price adjustment. The coefficient on the interaction term is negative and highly significant, which suggests that the lead-lag effect weakens as industry average market capitalization increases.

Panels B-D replace industry average size in the interaction term with industry average analyst coverage, institutional ownership, and turnover. In Section 4 I show that firms with higher levels of analyst coverage, institutional ownership, and trading volume lead other firms within the industry. The coefficients on all three interaction terms are negative and statistically significant at the 1% level, suggesting that the lead-lag effect is stronger in industries that have less analyst and institutional following,

Table 4
Cross-industry variation in lead-lag effect

LHS	$R_{i,1}(t-1)$	$R_{i,3}(t-1)$	$R_{i,3}(t-1)\times$ IC
	Panel A: IC = Average ln(Size)		
$R_{i,1}(t)$	0.21	0.40	−0.05
	*27.09******	*14.09******	*−7.92******
$R_{i,3}(t)$	−0.00	0.14	—
	−0.42	*17.80******	
	Panel B: IC = Average ln(# Analyst)		
$R_{i,1}(t)$	0.24	0.35	−0.14
	*26.99******	*10.24******	*−5.95******
$R_{i,3}(t)$	−0.02	0.13	—
	*−1.62******	*13.23******	
	Panel C: IC = Average ln(Institutional ownership)		
$R_{i,1}(t)$	0.23	0.23	−0.47
	*20.09******	*6.80******	*−5.05******
$R_{i,3}(t)$	−0.02	0.12	—
	*−1.86**	*10.48******	
	Panel D: IC = Average ln(Turnover)		
$R_{i,1}(t)$	0.21	0.10	−0.09
	*27.83******	*3.06******	*−5.91******
$R_{i,3}(t)$	−0.00	0.14	—
	−0.42	*17.80******	
	Panel E: IC = Average ln(Analyst Dispersion)		
$R_{i,1}(t)$	0.25	0.41	0.07
	*27.41******	*6.23******	*4.14******
$R_{i,3}(t)$	−0.02	0.13	—
	*−1.62**	*13.23******	
	Panel F: IC = HERF		
$R_{i,1}(t)$	0.21	0.19	0.16
	*28.21******	*21.92******	*3.02******
$R_{i,3}(t)$	−0.0031	0.1414	—
	−0.42	*17.80******	
	Panel G: IC = Five-firm ratio		
$R_{i,1}(t)$	0.20	0.16	0.17
	*24.10******	*10.95******	*3.81******
$R_{i,3}(t)$	−0.0031	0.1414	—
	−0.42	*17.80******	

I estimate the following regressions jointly across all industries, using weekly returns on the size portfolios formed within each industry:

$$R_{i,1}(t) = a_{i,0} + a_1 R_{i,1}(t-1) + b_1 R_{i,3}(t-1) + b_2 R_{i,3}(t-1) \times IC_i + e_{i,1}(t), \tag{T.4.1}$$

$$R_{i,3}(t) = c_{i,0} + c_1 R_{i,1}(t-1) + d_1 R_{i,3}(t-1) + e_{i,3}(t). \tag{T.4.2}$$

In Equations (T.4.1) and (T.4.2), $R_{i,1}(t)$ is the week t return on the portfolio of the smallest 30% firms from industry $i.R_{i,3}(t)$ is the week t return on the portfolio of the largest 30% firms from industry i. IC_i denotes the characteristic of industry i. Size is the market capitalization of a firm at the end of June of year t. It is matched with weekly returns from July of year t to June of year $t+1$. Analyst coverage is the average number of investment analysts following a firm from July of year $t-1$ to June of year t. Institutional ownership is measured in December of year $t-1$. Turnover is the average share turnover per week, which I define as the ratio of the number of shares traded in a week to the number of shares outstanding at the end of the week, averaged from July of year $t-1$ to June of year t. I double the turnover for NYSE/Amex firms so that it is comparable with Nasdaq turnover. I define analyst dispersion as the standard deviation of analysts' annual earnings forecast divided by the absolute value of the mean forecast each month averaged over the previous year. HERF refers to the Herfindhal index of an industry. I measure HERF by the sum of squared market share across all firms in an industry. The five-firm ratio is the combined market share of the top five firms in an industry. *Italics* indicate the t-statistics for the coefficient estimates. ***, **, and * denote significance at the 1, 5, and 10% levels, respectively.

and lower trading volume. These findings further support the lagged information diffusion hypothesis, because it is precisely in these less visible and neglected industries that I expect information to transmit more slowly.

I also investigate the impact of industry-level information uncertainty on the lead-lag effect within an industry. *A priori*, industries whose information environment is noisier should see slower information diffusion and a more pronounced lead-lag effect. Thus, in Panel E of Table 4, I measure industry-level information uncertainty with the average analyst dispersion, and interact it with the lagged return on big firms in the VARs. The coefficient on the interaction term is positive with a t-statistic of 4.14, which confirms my prediction that the lead-lag effect becomes stronger as information uncertainty increases.

The structure of the product markets may also affect the way news travels within an industry. In a highly concentrated industry in which the market is dominated by a few large firms, the dramatic difference in market power between big and small firms makes it more difficult for investors in small firms to determine the value implications of information generated on big firms. As a result, news diffuses slowly from big firms to small firms. On the other hand, in a highly competitive industry in which many competing firms share the market, news dissemination should be less sluggish. It should take investors in small firms less time and effort to extract the relevant information from past prices of big firms, since it is more likely that the information generated on big firms also has value implications for the industry as a whole.

As in Hou and Robinson (2006), I use two variables to measure the degree of competitiveness of an industry's product markets, the Herfindahl index (the sum of the squared market share across all firms in an industry), and the five-firm ratio (the combined market share of the top five firms in an industry). For both measures, a large value suggests that the industry is concentrated and a smaller value suggests that the industry is more competitive.

Panels F and G of Table 4 interact the two industry concentration measures with the lead-lag effect. The results support my conjecture that the lead-lag effect is stronger in concentrated industries, as evidenced by the positive, statistically significant coefficients on the interaction terms.

6. Lead-lag Effect and News in Earnings Announcements

I now take a slightly different approach by focusing on news in earnings announcements, since earnings news, apart from past price movements, is another source of information that investors often use to assess the future prospects of their stocks. First, I use the event study method to examine the impact of earnings surprises of big firms on the stock prices of small firms from the same industry. According to the slow information diffusion hypothesis, the stock prices of small firms should react positively but with a delay when big firms from the same industry announce unexpectedly strong earnings, and negatively when they announce unexpectedly poor earnings. Once established, I then relate the sluggish adjustment of small firms to past earnings surprises on big firms to the lead-lag effect in returns using Fama and MacBeth (1973) style cross-sectional regressions.

6.1 Intra-industry event study of earnings announcements

I use two variables to measure earnings surprises. My first measure is the abnormal return relative to an equal-weighted benchmark portfolio of firms matched by size, BE/ME, and past 1-year return, for the week surrounding an earnings announcement. This variable does not require an earnings expectation model. Instead, it relies on the market's reaction to gauge the surprise element in an earnings announcement. My second measure is the standardized unexpected earnings (SUE), which is the change in current quarter's earnings from four quarters ago, divided by the standard deviation of the unexpected earnings over the last eight quarters.

For each industry, I compute the average earnings surprise of the largest 30% firms and rank the announcement weeks according to it. I define the weeks for which the average surprise on big firms is above the 90th percentile breakpoint as intra-industry positive earnings events and those below the 10th percentile breakpoint as negative earnings events. I then compute the average abnormal returns of the smallest 30% firms within the same industry from 4 weeks before to 12 weeks after positive and negative earnings events, using the event study method of Jaffe (1974) and Mandelker (1974). For each calendar week t, I calculate the abnormal return of each small firm that has a qualified intra-industry earnings event in week $t - k$ ($k = -4$ to 12), by benchmarking against an equal-weighted portfolio matched by size, BE/ME, and past 1-year return. I average the abnormal returns across firms (and industries) for each calendar week and then across time to compute the average abnormal return for event week k ($k = -4$ to 12). This procedure takes into account the correlation of abnormal returns across event firms, so that appropriate inferences can be drawn on the magnitude of these abnormal returns.

I perform the calculation separately for positive and negative earnings events, and report the average abnormal returns and their t-statistics in Table 5. The results using the return-based earnings surprise measure is similar to those using SUE. To conserve space, I report only the results for the return-based earnings surprise measure.

First, consistent with the findings in accounting studies on the information transfer effect [Foster (1981), Han and Wild (1990) and Schipper (1990)], I find a positive relation between the big firms' earnings surprises and the contemporaneous price reactions of small firms from

Table 5
Intra-industry event study of earnings announcements

	Positive earnings events		Negative earnings events	
Event week	Abnormal return	*t-stat*	Abnormal return	*t-stat*
−4	0.0005	*0.78*	−0.0006	*-0.90*
−3	0.0006	*1.08*	−0.0014	*-2.55*
−2	0.0003	*0.45*	−0.0003	*-0.46*
−1	0.0022	*3.42*	−0.0006	*-0.77*
0	0.0044	*5.72*	−0.0028	*-3.93*
1	0.0050	*6.37*	−0.0033	*-5.05*
2	0.0022	*3.92*	−0.0016	*-2.58*
3	0.0028	*3.98*	−0.0016	*-2.54*
4	0.0020	*2.87*	−0.0020	*-3.49*
5	0.0005	*0.88*	−0.0014	*-2.18*
6	0.0013	*1.80*	−0.0014	*-2.00*
7	0.0017	*2.79*	−0.0010	*-1.31*
8	0.0014	*2.28*	−0.0008	*-1.04*
9	0.0006	*1.04*	−0.0004	*-0.71*
10	0.0011	*1.93*	−0.0008	*-1.41*
11	0.0004	*0.69*	−0.0010	*-1.72*
12	0.0014	*2.17*	−0.0008	*-1.35*
(1, 4)	0.0023	*6.83*	−0.0022	*-6.11*

Table 5 reports the average abnormal returns of the smallest 30% firms in an industry from 4 weeks before to 12 weeks after big firms within the industry announce unexpected strong (weak) earnings. Within each industry, for the weeks surrounding their quarterly earnings announcements I calculate the average abnormal returns of the largest 30% firms relative to a characteristic-based benchmark portfolio matched by size, BE/ME, and past 1-year return. I define the weeks for which the average abnormal return is above the 90th percentile breakpoint as positive earnings events and those below the 10th percentile breakpoint as negative earnings events. For each calendar week t, I calculate the abnormal return on each small firm that has a qualified intra-industry earnings event in week $t - k$ ($k = -4$ to 12) by benchmarking against an equal-weighted portfolio of firms matched by size, BE/ME and past 1-year return. To compute the average abnormal return of small firms for event week k, I average the individual abnormal returns across firms for each calendar week t and then across time. I also report the average weekly abnormal return and t-statistic for small firms over the 4 weeks following an event.

the same industry. On average, small firms experience an abnormal return of 44 (−28) basis points with a t-statistic of 5.72 (−3.93) when big firms within the industry announce unexpected strong (poor) earnings. But more important, small firms' stock prices continue to drift upward (downward) following positive (negative) earnings shocks to big firms. The average weekly abnormal return for small firms is 23 basis points (t-statistic 6.83) over the 4 weeks following positive earnings surprises on big firms and−22 basis points (t-statistic−6.11) following negative surprises. Figure 1 plots the cumulative abnormal returns of small firms from week−4 to week 12 in event time. The post-announcement drift of small firms is also evident from the figure.

6.2 Regression analysis of intra-industry lead-lag effect and post-earnings announcement drift

I want to know if the lead-lag effect in returns is related to the post-announcement drift of small firms following the earnings releases of big firms within the same industry. Therefore, for every calendar week, I

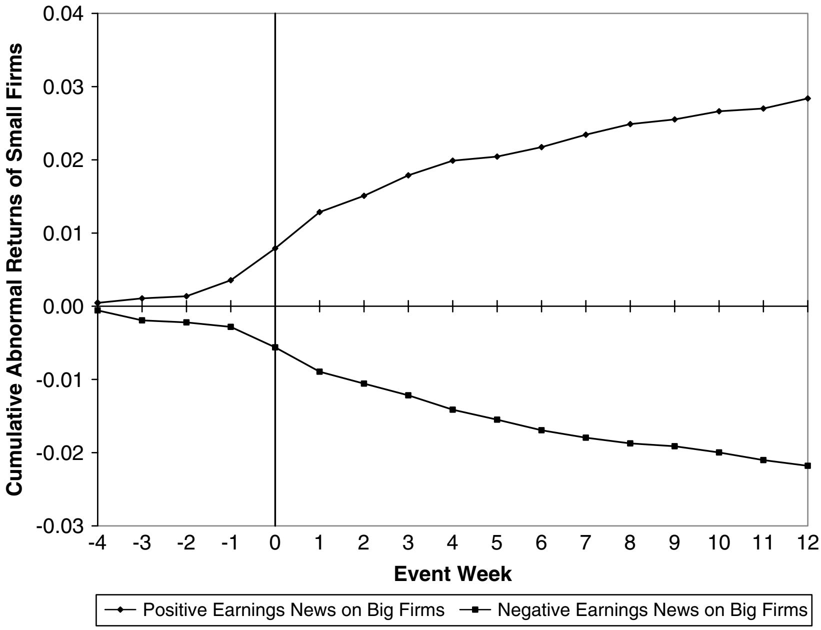

Figure 1
Intra-industry event study of earnings announcements
The figure plots in event time the cumulative average abnormal returns of the smallest 30% firms within an industry from 4 weeks before to 12 weeks after big firms within the same industry announce unexpected strong (weak) earnings. Within each industry, I calculate the average abnormal returns of the largest 30% firms relative to a characteristic benchmark portfolio matched by size, BE/ME, and past 1-year return for the weeks surrounding their quarterly earnings announcements. I define the weeks for which the average abnormal return is above the 90th percentile breakpoint as positive earnings events and those below the 10th percentile breakpoint as negative earnings events. For each calendar week t, I calculate the abnormal return on each small firm that has a qualified intra-industry earnings event in week $t - k$ ($k = -4$ to 12) by benchmarking against an equal-weighted portfolio of firms matched by size, BE/ME, and past 1-year return. To compute the average abnormal return of small firms for event week k, I average the individual abnormal returns across firms for each calendar week and then across time.

estimate a cross-sectional regression on the smallest 70% firms from every industry. I regress weekly individual stock returns on size, BE/ME, a small-firm's own earnings surprise as of the most recent quarterly announcement, earnings surprises on big firms (the average earnings surprise for the largest 30% firms in the industry to which a small firm belongs), and lagged returns on big firms (weekly returns from week $t-4$ to week $t-1$ on the portfolio of the largest 30% firms in the industry to which a small firm belongs). Table 6 reports the time-series averages of the weekly cross-sectional regression coefficients and their time-series t-statistics.

I use the same abnormal return-based measure for earnings news as in the earnings event study. (I also use SUE to measure earnings news and obtained nearly identical results.) The regression results show that a small firm's own earnings surprise (AR) is positively and significantly correlated

Table 6
Fama and MacBeth (1973) cross-sectional regressions: intra-industry lead-lag effect and past earnings surprises

ln(Size)	ln(BE/ME)	AR	$AR_{i,3}$	$R_{i,3}(-1)$	$R_{i,3}(-2)$	$R_{i,3}(-3)$	$R_{i,3}(-4)$
−0.0009	0.0006						
−6.95	*4.35*						
−0.0008	0.0006	0.0055					
−6.87	*4.30*	*6.09*					
−0.0008	0.0006		0.1266				
−6.82	*4.70*		*8.91*				
−0.0008	0.0006	0.0054	0.1257				
−6.74	*4.65*	*5.94*	*8.86*				
−0.0009	0.0006	0.0050	−0.0013	0.1684	0.0721	0.0781	0.0105
−6.71	*5.72*	*5.61*	*−0.04*	*11.00*	*2.35*	*4.86*	*0.64*

For each week from January 1972 to December 2001, I estimate firm-level cross-sectional regressions of weekly returns of the smallest 70% firms from each industry on earnings surprises and past returns of big firms from the same industry. $R_{i,3}(-1)$, $R_{i,3}(-2)$, $R_{i,3}(-3)$, and $R_{i,3}(-4)$ are the past 4 weeks' returns on the portfolio of the largest 30% firms from the industry to which each small firm belongs. I measure a firm's earnings surprise by AR, the abnormal return, relative to an equal-weighted benchmark portfolio of firms matched by size, BE/ME and past 1-year return, for the week surrounding its most recent quarterly earnings announcement. $AR_{i,3}$ is the average AR of the largest 30% firms from the industry to which each small firm belongs. I report the time-series averages of the coefficients from the weekly regressions. *Italics* indicate their time-series t-statistics.

with future returns. The average earnings surprise on big firms ($AR_{i,3}$) enters the regression positively with a t-statistic of 8.91, which confirms the finding from the event study that small firms experience positive drift after big firms within the industry announce unexpected strong earnings and negative drift after they announce unexpected poor earnings. $AR_{i,3}$ remains significant even after I include a small firm's own earnings surprise (AR) in the regression, suggesting that the earnings surprises on big firms add information about the future prospects of small firms that is not contained in small firms' own earnings surprises.[5]

In the last regression specification of Table 6, I include the past 4 weeks' returns on big firms [$R_{i,3}(-k)$, $k = 1$ to 4] as additional independent variables. Doing so allows me to separate the predictive power of past returns of big firms from that of past earnings surprises of big firms. The regression shows that once I account for the lead-lag effect in returns, the average earnings surprise on big firms loses its predictive power for future small-firm returns. This result suggests that the lead-lag effect in returns is related to the tendency of small-firm stock prices to react slowly to news in past earnings announcements on big firms, and, as far as forecasting future returns on small firms is concerned, past returns on big firms contain less noisy signals than do past earnings surprises. Therefore,

[5] It also appears that the returns on small firms respond more strongly to earnings surprises on big firms than to their own earnings surprises. In the cross-sectional regressions, when I standardize AR and $AR_{i,3}$ to have zero mean and unit variance to make them comparable and then rerun the regressions, both the point estimate and the t-statistic for $AR_{i,3}$ are greater than are those for AR.

the evidence in Table 6 provides additional support to the hypothesis that slow information diffusion drives the lead-lag effect in stock returns.

7. Conclusion

In this article, I find that the lead-lag effect first documented by Lo and MacKinlay (1990) contains a persistent and highly significant industry component. Big firms lead small firms within the same industry. This intra-industry lead-lag effect drives the overall lead-lag effect. Moreover, this effect is primarily driven by small firms' sluggish response to negative news on big firms.

Other known determinants of the lead-lag effect (e.g., analyst coverage, institutional ownership, and trading volume) remain significant intra-industry. Furthermore, firms with high industry market share lead firms with low market share, and firms with lower analyst dispersion lead their industry peers with higher analyst dispersion. The lead-lag effect is stronger in small, neglected, and concentrated industries.

I also show that the lead-lag effect is related to the post-announcement drift of small firms following earnings releases of big firms from the same industry.

My findings cannot be explained by nonsynchronous trading or time-varying expected returns, and are mainly consistent with the hypothesis that information diffuses slowly across firms, and that industries are the primary channel for news dissemination in the equity markets. More work on measuring and distinguishing the impacts of various impediments to the information diffusion process could lead to richer predictions and empirical tests. I leave them for future research.

References

Badrinath, S. G., J. R. Kale, and T. H. Noe, 1995, "Of Shepherds, Sheep, and the Cross-Autocorrelations in Equity Returns," *Review of Financial Studies*, 8, 401–430.

Barron, O., O. Kim, S. Lim, and D. Stevens, 1998, "Using Analysts' Forecasts to Measure Properties of Analysts' Information Environment," *Accounting Review*, 73, 421–433.

Barron, O., and P. Stuerke, 1998, "Dispersion in Analyst's Earnings Forecasts as a Measure of Uncertainty," *Journal of Accounting, Auditing, and Finance*, 13, 243–268.

Boudoukh, J., M. P. Richardson, and R. F. Whitelaw, 1994, "A Tale of Three Schools: Insights on Autocorrelations of Short-Horizon Stock Returns," *Review of Financial Studies*, 7, 539–573.

Brennan, M. J., N. Jagadeesh, and B. Swaminathan, 1993, "Investment Analysis and the Adjustment of Stock Prices to Common Information," *Review of Financial Studies*, 6, 799–824.

Chordia, T., and B. Swaminathan, 2000, "Trading Volume and Cross-Autocorrelations in Stock Returns," *Journal of Finance*, 55, 913–935.

Conrad, J., and G. Kaul, 1988, "Time Varying Expected Returns," *Journal of Business*, 61, 409–425.

Conrad, J., and G. Kaul, 1989, "Mean-Reversion in Short-Horizon Individual Security Return," *Journal of Financial Economics*, 29, 365–384.

Diamond, D., and R. Verrecchia, 1987, "Constraints on Short-Selling and Asset Price Adjustment to Private Information," *Journal of Financial Economics*, 18, 277–311.

Fama, E. F., and J. MacBeth, 1973, "Risk, Return and Equilibrium: Empirical Tests," *Journal of Political Economy*, 81, 607–636.

Foster, G., 1981, "Intra-Industry Information Transfers Associated with Earnings Release," *Journal of Accounting and Economics*, 3, 201–232.

Hameed, A., 1997, "Time-Varying Factors and Cross-Autocorrelations in Short-Horizon Stock Returns," *Journal of Financial Research*, 20, 435–458.

Han, J. C. Y., and J. J. Wild, 1990, "Unexpected Earnings and Intra-Industry Information Transfers: Further Evidence," *Journal of Accounting Research*, 28, 211–219.

Hou, K., and T. J. Moskowitz, 2005, "Market Frictions, Price Delay, and the Cross-Section of Expected Returns," *Review of Financial Studies*, 18, 981–1020.

Hou, K., and D. T. Robinson, 2006, "Industry Concentration and Average Stock Returns," *Journal of Finance*, 61, 1927–1956.

Imhoff, E., and G. Lobo, 1992, "The Effect of Ex Ante Earnings Uncertainty on Earnings Response Coefficients," *Accounting Review*, 67, 427–439.

Jaffe, J. F., 1974, "Special Information and Insider Trading," *Journal of Business*, 47, 410–428.

LaPlante, M., and C. J. Muscarella, 1997, "Do Institutions Receive Comparable Execution in the NYSE and Nasdaq Markets? A Transaction Study of Block Trades," *Journal of Financial Economics*, 45, 97–134.

Lo, A. W., and A. C. MacKinlay, 1990, "When Are Contrarian Profits Due to Stock Market Overreaction?" *Review of Financial Studies*, 3, 175–205.

Loughran, T., and P. Schultz, 2005, "Liquidity: Urban Versus Rural Firms," *Journal of Financial Economics*, 78, 341–374.

Mandelker, G., 1974, "Risk and Return: The Case of Merging Firms," *Journal of Financial Economics*, 1, 303–336.

McQueen, G., M. Pinegar, and S. Thorley, 1996, "Delayed Reaction to Good news and the Cross-Autocorrelation of Portfolio Returns," *Journal of Finance*, 51, 889–919.

Mech, T. S., 1993, "Portfolio Return Autocorrelation," *Journal of Financial Economics*, 34, 307–344.

Schipper, K., 1990, "Information Transfers," *Accounting Horizons*, 4, 97–107.

Zhang, F. X., 2006, "Information Uncertainty and Stock Returns," *Journal of Finance*, 61, 105–137.