JOURNAL OF FINANCIAL AND QUANTITATIVE ANALYSIS Vol. 53, No. 5, Oct. 2018, pp. 1937–1961

doi:10.1017/S0022109018000261

Know Thy Neighbor: Industry Clusters, Information Spillovers, and Market Efficiency

Joseph Engelberg, Arzu Ozoguz, and Sean Wang*

Abstract

Firms in industry clusters have market prices that are more efficient than firms outside clusters. To establish causality, we analyze exogenous firm relocations and find that firms that relocate into industry clusters have higher levels of industry information in their prices. We argue that geographical proximity allows for information spillovers, reducing marginal cost to information producers. Our evidence supports this view: Analysts are more likely to cover stocks inside industry clusters, and when institutional investors have a large position in one stock in the industry cluster, they are more likely to hold other stocks in the same industry cluster.

> After all, geographical proximity matters in transmitting knowledge because … intellectual breakthroughs must cross hallways and streets more easily than oceans and continents.
>
> —Audrestsch and Feldman ((2004), p. 88)

I. Introduction

Prices reflect information. When information is costly, the amount of information impounded in price will directly reflect the cost of information (Grossman and Stiglitz (1980)) and investors' choices on which assets to learn about (Veldkamp (2006)).

Recently, many articles have provided evidence that locality reduces the cost of information. The idea is simple: A fund manager located in San Diego will be able to collect information about each biotech firm in San Diego more easily

*Engelberg, jengelberg@ucsd.edu, University of California at San Diego Rady School of Management; Ozoguz (corresponding author), aozoguz@smu.edu, Southern Methodist University Cox School of Business; and Wang, sean_wang@kenan-flagler.unc.edu, University of North Carolina at Chapel Hill Kenan-Flagler Business School. We have benefited from discussions with an anonymous referee, Hendrik Bessembinder (the editor), Robert Bushman, Chris Malloy, Chris Parsons, and seminar participants at the University of Southern California, Rutgers University, the University of Washington, the University of Texas at Dallas, the 2010 Financial Intermediation Research Society Conference, the 2010 Annual Asset Pricing Retreat, and the 2011 American Finance Association Meeting.

than a fund manager in Chicago. Among other things, proximity will allow the fund manager easier access to management and information about local inputs to production. It is now well established that the *geographic proximity between firms and investors* can influence both the behavior of information producers and asset prices (Coval and Moskowitz (1999), (2001), Loughran and Schultz (2005)).

This article suggests a new channel by which geography can affect asset prices: the *proximity between one firm and its industry peers.* Our central tenet is that geographic proximity to other peer firms should improve a firm's information environment for investors by lowering their marginal cost of collecting information. Consider again the Chicago fund manager who collects information about one biotech firm in San Diego. If biotech firms in San Diego share information and are exposed to common, local inputs to production, then the Chicago manager can use his or her information about the first biotech firm in San Diego to better understand the second biotech firm in San Diego at a lower cost.

Because we do not directly observe the marginal cost of acquiring information on peer firms, we use the amount of information impounded in prices as a proxy for its cost and test whether stock prices of firms located proximally to other peer firms impound information more quickly. We find evidence that suggests they do. Using data on all publicly traded U.S. stocks between 1990 and 2007 and the price delay measure of Hou and Moskowitz (2005), we find that prices respond more (less) quickly to industry information for firms located inside (outside) industry clusters. Moving into an industry cluster improves the median firm's price delay measure by more than 10% by itself; when we control for other characteristics, the magnitude of the cluster effect on price delay still remains large and significant (approximately equal to half of that of the effect of market capitalization).

Of course, being inside an industry cluster is not a random event. To address endogeneity concerns, we also study a special set of firms that relocate. Using Compact Disclosure to identify all firm relocations between 1990 and 2006, we hand-collect and identify moves that are neither related to a change in the firm's business or strategy nor driven by the firm's fundamentals in a way that might influence information production outside the geography channel. For example, we remove the relocation of AppliedMicro from San Diego to Sunnyvale in 2005 because it was part of AppliedMicro's acquisition of 3ware, which was located in Sunnyvale. We select only those relocations that appear "exogenous" to all other factors, such as Verilink's move from San Jose to Huntsville in order to reduce operating costs, Fair Isaac's relocation from San Jose to Minneapolis to be closer to executives' homes in Minnesota, and the move of Trico Marine Services from New Orleans to Houston following Hurricane Katrina. Our pre-post analyses of these relocating firms reveal that firms that move into an industry cluster have prices that subsequently impound industry information more quickly. Although, admittedly, the size of our relocation sample is quite small to draw definitive conclusions, we hope that our results from these analyses provide some insight to the discerning reader. To the best of our knowledge, this is the first article to find evidence that geographic proximity among peer firms affects the incorporation of information into prices.

Given our main finding that stock prices of firms located in industry clusters reflect more information, we next investigate the behavior of information producers to understand why this might be. Here we find circumstantial evidence that producers of information (i.e., analysts and institutional investors) use information about one firm in a cluster to learn about another. Specifically, we find that i) financial analysts are more likely to cover firms that are located within a more concentrated industry cluster, and ii) institutional investors are more likely to take large positions in these firms.[1] These results remain robust even after controlling for an array of explanatory variables, as well as year, industry, and Metropolitan Statistical Area fixed effects.

We also aggregate the holdings of institutional investors' portfolios to examine their holdings *within* industry clusters. We find that fund managers who tilt their portfolios toward large firms within a given industry cluster also tend to hold a larger number of smaller-sized firms *within the same industry cluster*. Following evidence from Lo and MacKinlay (1990) and Hou (2007) that information diffuses from larger firms to smaller firms, we interpret this evidence to be consistent with a manager's choice to optimize the cost of gathering information by first learning about a given firm within an industry cluster and then applying the correlated information to other firms within the same locality.

Finally, we consider alternative explanations for why stock prices of firms inside clusters might incorporate more industry information. One possible explanation is that local fund managers are more likely to hold and be informed about local stocks (Coval and Moskowitz (1999), (2001)). However, when we exclude local fund managers from our sample, our results remain unchanged.

Another possibility is that geography represents a finer classification of industry. That is, firms inside an industry cluster may be more likely to be in the same line of business. If industry classifications by Standard Industrial Classification (SIC) codes are noisy and two firms inside an industry cluster are more likely to be truly in the same kind of business than two firms outside an industry cluster, this might explain why fund managers and analysts who might specialize in a particular industry focus on industry clusters. In order to address this possibility, we look at the differences between the 3-digit and 4-digit SIC codes of firms located in and out of industry clusters. We find no evidence that geographically proximate firms might proxy for a finer classification of industry.

Taken together, our findings expand the role of geography in setting prices. Whereas the well-documented home bias among institutional investors suggests that locality reduces the cost of information *only for local funds*, our study suggests that spillovers among local peer firms reduce the cost of information *for all funds*. This alternative channel implies that the relationship between geography and asset pricing may be more general than previously thought.

The article is organized as follows: In Section II, we describe the data and the construction of variables. In Section III, we examine the relationship between firm

[1]This finding is also consistent with Hameed, Morck, Shen, and Yeung (2015), who show that more analysts cover firms whose fundamentals are good predictors of many other firms' fundamentals.

location and stock price delay. In Section IV, we consider a special group of exogenous firm relocations to address endogeneity issues. In Section V, we present the information-spillover hypothesis as a plausible explanation for our main findings and provide evidence from information producers that is consistent with this hypothesis. Section VI discusses alternative hypotheses. Section VII concludes.

II. Data and Variable Construction

Our sample includes all securities listed on the New York Stock Exchange (NYSE), American Stock Exchange (AMEX), and National Association of Securities Dealers Automated Quotations (NASDAQ) with share codes 10 or 11 that have available data from Center for Research in Security Prices (CRSP) and Compustat databases between 1990 and 2007. We merge prices from CRSP with accounting data from Compustat annual files, using the CRSP Link database. In addition, we require each firm to have i) a nonmissing SIC code in CRSP and ii) state and county codes associated with the company headquarters from Compustat annual files. We exclude firms located in Hawaii and Puerto Rico, financial firms (SICs 6000–6999), regulated utilities (SICs 4000–4999), and any other 3-digit SIC industries that have less than 10 firms in any of the sample years. All accounting variables are winsorized at the 1% level.

Following Coval and Moskowitz (1999), Pirinsky and Wang (2006), and Almazan, Motta, Titman, and Uysal (2010), we define each firm's locality as the geographical location of its headquarters. We use Metropolitan Statistical Areas (MSAs) as defined by the 1990 U.S. Census Bureau to proxy for geographical location[2] and use state and county code classifications from Compustat annual files to merge our sample firms with the MSA codes.

Although one concern with the Compustat location data is that Compustat only reports the *current* state and county of firms' headquarters,[3] in Section IV, we specifically examine a subset of firms in our sample that relocate during our sample period. We merge our data with the Compact Disclosure database, which provides the zip code of the firm's headquarters on an annual basis. Using Compact Disclosure, we first annually map the zip codes of firms' headquarters into MSAs and then identify all firms whose headquarters have moved from one MSA to another. Unfortunately, the data availability for Compact Disclosure constrains our sample period to 1990–2016 for this analysis.

[2]The general concept of an MSA, as defined by the U.S. Census Bureau, is that of a geographical region with a large population nucleus, together with adjacent communities having a high degree of social and economic integration with that core. Where appropriate, we replace the MSA with the broader Consolidated Metropolitan Statistical Area (CMSA) definition, which groups together a number of adjacent MSAs. A CMSA has a population of 1 million or more and consists of separate components that are themselves considered Primary Metropolitan Statistical Areas (PMSAs). For example, the San Francisco–Oakland–San Jose CMSA in California consists of 6 PMSAs (Oakland, San Francisco, San Jose, Santa Cruz, Santa-Rosa–Petaluma, and Napa). Our use of CMSAs allows us to make sure we account for clusters that can extend beyond MSAs.

[3]Pirinsky and Wang (2006) show that in the period 1992–1997, less than 2.4% of firms in Compustat changed headquarters locations.

Throughout the article, we use 3-digit SIC codes. This choice reflects a balance between minimizing the possibility of grouping unrelated firms together while ensuring the viability of an "industry cluster" definition. We measure industry clustering using a continuous variable, CLUSTER_RATIO, which we compute as the number of firms in a given industry within the same MSA code scaled by the total number of firms within the same industry. By construction, higher levels of CLUSTER_RATIO indicate more concentrated industry clustering. We also identify clustered-firms using a binary variable, CLUSTER_DUMMY, which takes a value of 1 if a firm's MSA includes 10 or more firms with the same 3-digit SIC, and 0 otherwise.

Our final sample includes 198 MSAs, 61 industries, 7,256 firms, and 52,697 firm-year observations. Based on the CLUSTER_DUMMY, 34% of firms are located inside an industry cluster. Our industry cluster measure, CLUSTER_RATIO, has a mean of 0.10, suggesting that, on average, 10% of firms within a 3-digit SIC code reside in the same MSA in our sample. This variable ranges from 0.001 to 0.529 in our data (unreported). This number is 15% among MSAs with clusters and drops to 6% when we condition on noncluster areas.

In Table 1, we compare firms located inside clusters with those that are located outside clusters, as identified by CLUSTER_DUMMY. We find that cluster

TABLE 1

Cluster/Out-of-Cluster Differences

Table 1 compares the mean values for firms that are outside clusters to those for firms that are located inside clusters. A firm is classified as located in a cluster if there are 10 or more firms with the same 3-digit Standard Industrial Classification (SIC) code located in the same Metropolitan Statistical Area (MSA). CLUSTER_RATIO is defined as the number of firms in a given industry, as defined by the 3-digit SIC code and the same MSA code, scaled by the total number of firms in the same industry. All financial variables and the NUMBER_OF_EMPLOYEES are taken from Compustat. MARKET_CAPITALIZATION is the market value of equity computed from Center for Research in Security Prices (CRSP) data as the share price times the number of shares outstanding on the fiscal year-end date. MARKET_TO_BOOK is the ratio of market value of equity to book equity. ROA is return on assets, measured as earnings before interest, taxes, and amortization scaled by the previous year's total assets. SALES_GROWTH is the annual percentage change in sales. TANGIBLE_ASSETS is net fixed assets, defined as plant, property, and equipment. CAPITAL_EXPENDITURE, CAPITAL_EXPENDITURE+R&D, TANGIBLE_ASSETS, and CASH are scaled by the previous year's total assets. All liquidity measures are obtained from CRSP. TRADING_VOLUME is the yearly average of monthly dollar trading volume. TURNOVER is the yearly average of the monthly number of shares traded divided by the number of shares outstanding. ILLIQUIDITY is the average daily absolute return over daily trading volume, constructed as in Amihud (2002). The *t*-statistics are clustered by firm. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Mean				
Variables	All Firms	Firms Outside Clusters	Firms Inside Clusters	Difference	*t*-Stat.
CLUSTER_RATIO	0.09	0.06	0.15	0.09***	36.97
ASSETS	341.77	362.68	301.92	−60.76***	−3.92
MARKET_CAPITALIZATION	1,723.56	1,490.66	2,175.61	684.95*	1.83
SALES	338.07	389.54	237.41	−152.13***	−9.96
CAPITAL_EXPENDITURE	0.08	0.076	0.073	−0.002	−0.89
CAPITAL_EXPENDITURE+R&D	0.19	0.153	0.253	0.10***	18.42
ROA	0.07	0.095	0.019	−0.08***	−11.49
SALES_GROWTH	0.25	0.222	0.319	0.10***	5.76
SGA/SALES	0.49	0.390	0.709	0.32***	15.42
TANGIBLE_ASSETS	0.26	0.289	0.219	−0.07***	−8.71
MARKET_TO_BOOK	2.62	2.310	3.238	0.93***	13.69
R&D	0.13	0.085	0.199	0.11***	21.54
ADV_EXP/SALES	0.06	0.047	0.082	0.04***	3.28
CASH	0.29	0.22	0.44	0.22***	24.11
NUMBER_OF_EMPLOYEES	6.08	7.54	3.23	−4.31***	−5.99
TRADING_VOLUME	4,184.13	2,656.85	7,881.35	5,224.50***	5.47
ILLIQUIDITY	0.40	0.49	0.19	−0.30***	−12.40
TURNOVER	1.58	1.31	2.25	0.94***	17.72

and noncluster firms differ across several dimensions. Firms inside clusters are slightly smaller in terms of assets and sales revenue, but they have significantly higher market cap, research and development (R&D), sales growth, cash balances, and market-to-book ratios and larger levels of intangibles in the form of advertising expenses and selling, general, and administrative (SGA) expenses. Firms inside clusters are also less profitable than those outside clusters. These findings are largely consistent with Almazan et al. (2010).

III. Price Informativeness among Firms in Clusters

We begin by asking whether firms located close to other peer firms, that is, firms that reside in more concentrated industry clusters, have higher levels of informational efficiency. Our question is motivated by prior literature that has shown evidence of correlated fundamentals and the presence of local information networks among colocated firms (Dougal, Parsons, and Titman (2015), Kedia and Rajgopal (2009), and Engelberg, Gao, and Parsons (2012)).[4] If firms in industry clusters have correlated fundamentals, then the marginal cost of learning for firms within an industry cluster should be lower than the cost for those outside of industry clusters because information producers can more directly apply information from one firm to others that reside in the same cluster.

Because we do not directly observe the marginal cost of information for each firm, we test this conjecture using the amount of information impounded in prices as a proxy for its cost. Specifically, we follow Hou and Moskowitz (2005) and construct a price-delay measure to capture the speed of price adjustment to common industry information.[5] To construct the industry price-delay measure, we first regress each individual stock's weekly returns on the contemporaneous and 4-week-lagged returns on the market and industry portfolios over the previous 3 years. Specifically, we estimate the following regression:

$$(1) \quad r_{i,t} = \alpha_i + + \beta_0 r_{M,t} + \sum_{n=1}^{4} \beta_n r_{M,t-n} + \delta_0 r_{\text{IND},t} + \sum_{n=1}^{4} \delta_n r_{\text{IND},t-n} + \epsilon_{i,t},$$

where $r_{i,t}$ is the weekly return on stock i, and $r_{M,t}$ and $r_{\text{IND},t}$ denote the weekly return on the market and the industry portfolio, respectively. To control for common market-wide information, we also include the contemporaneous and lagged market returns in equation (1).

[4]There is a large literature on urban agglomeration that demonstrates knowledge spillovers within geographic clusters (e.g., Jacobs (1969)), arguing that geographic clusters spur information creation, dissemination, and learning. Christ (2009) surveys the literature and catalogues 61 articles that identify knowledge spillovers through channels such as employment, productivity, and patent activity. Most recently, Ellison, Glaeser, and Kerr (2010) find evidence that industry clusters exist not only to save transport and labor costs but also to benefit from "intellectual spillovers."

[5]Hou and Moskowitz (2005) create a price delay measure that examines the relation between individual stock returns and lagged market returns to proxy for the rate at which market-wide information is incorporated in the price of an individual stock. See Engelberg, Gao, and Jagannathan (2008) for a similar construction of industry information delay measure.

After estimating the coefficients in equation (1), we next identify the delay with which a stock price responds to industry-wide information by constructing three versions of a price-delay measure that captures the following intuition: if the stock responds to industry-wide news immediately, then δ_0 should be significantly different from 0, but the lagged coefficients δ_n should not be different from 0. The first measure, IND1, formalizes this intuition by measuring the fraction of variation in contemporaneous individual stock returns explained by lagged industry returns. That is, 1 minus the ratio of R^2 from the regression of equation (1), where $\delta_n, \forall n \in [1,4]$ are restricted to 0, scaled by the R^2 from the regression of equation (1) with no restrictions:

$$\text{(2)} \qquad \text{IND1} = 1 - \frac{R^2_{\delta_n=0,\forall n\in[1,4]}}{R^2}.$$

Larger values for IND1 indicate that more return variation is captured by lagged industry returns and hence suggest a slower speed of industry-wide information diffusion.

In addition, we construct two alternative measures, IND2 and IND3, which incorporate the economic magnitude and the precision of the lagged estimates:

$$\text{(3)} \qquad \text{IND2} = \frac{\sum_{n=1}^{n=4} n\delta_n}{\delta_0 + \sum_{n=1}^{n=4} \delta_n}$$

and

$$\text{(4)} \qquad \text{IND3} = \frac{\sum_{n=1}^{n=4} n\left[\frac{\delta_n}{\text{se}(\delta_n)}\right]}{\frac{\delta_0}{\text{se}(\delta_0)} + \sum_{n=1}^{n=4}\left[\frac{\delta_n}{\text{se}(\delta_n)}\right]},$$

where $\text{se}(\cdot)$ denotes the standard error of the coefficient estimate.[6] Similarly, IND2 and IND3 also increase with slower diffusion of industry information.

Finally, we test our main conjecture that firms in industry clusters incorporate industry information more quickly. Specifically, using a linear regression model of industry price delay, we test whether firms' price-delay measures are systematically related to industry clustering concentrations as measured by CLUSTER_RATIO.

Table 2 presents regression results with an array of controls. In all specifications, we correct standard errors for within-industry clustering. In column 1, we

[6]Variations of these measures have also been employed by Brennan, Jegadeesh, and Swaminathan (1993) and Mech (1993).

TABLE 2
Industry Clusters and Price Delay

In Table 2, IND1 is the industry-wide price-delay measure constructed as in equation (2) in the text. CLUSTER_RATIO is defined as the number of firms in a given industry, as defined by the 3-digit Standard Industrial Classification (SIC) code, and in the same MSA code, scaled by the total number of firms in the same industry. LOG_SIZE is the natural logarithm of firm size, measured as market capitalization. LOG_NUM_EST is the natural logarithm of 1 plus the amount of analyst coverage for a firm in a given year, measured as the aggregate number of year-ahead earnings estimates as found in the Institutional Brokers' Estimate System (IBES) Detail History file. LOG_NUM_INST is the natural logarithm of the average of the total number of institutions currently holding the stock over the calendar year. LOG_VOL is the natural logarithm of the yearly average of monthly dollar trading volume. LOG_TURNOVER is the natural logarithm of the yearly average of the monthly number of shares traded divided by the number of shares outstanding. LOG_ILLIQ is the natural logarithm of the average daily absolute return over daily trading volume, constructed as in Amihud (2002). Standard errors are clustered by industry and are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: IND1							
Variable	1	2	3	4	5	6	7	8
CLUSTER_RATIO	−0.212***	−0.161***	−0.059***	−0.047***	−0.043***	−0.029**	−0.038***	−0.043**
	(0.0229)	(0.0264)	(0.0130)	(0.0097)	(0.0096)	(0.0105)	(0.021)	(0.019)
LOG_SIZE			−0.041***	−0.025***	−0.007***	0.004*	0.004*	0.005*
			(0.0013)	(0.0013)	(0.0016)	(0.0021)	(0.0021)	(0.002)
LOG_NUM_EST				−0.042***	−0.025***	−0.019***	−0.019***	−0.036***
				(0.0028)	(0.0028)	(0.0027)	(0.0027)	(0.003)
LOG_NUM_INST					−0.038***	−0.032***	−0.032***	−0.032***
					(0.0017)	(0.0018)	(0.0019)	(0.0018)
LOG_VOL						−0.007**	−0.008***	−0.007***
						(0.0032)	(0.0033)	(0.002)
LOG_TURNOVER						0.005	0.005	0.005
						(0.0032)	(0.0033)	(0.0032)
LOG_ILLIQ						0.010***	0.010***	0.010***
						(0.0013)	(0.0013)	(0.0013)
Year fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA fixed effects	No	No	No	No	No	No	Yes	No
Firm fixed effects	No	No	No	No	No	No	No	Yes
No. of obs.	49,759	49,759	49,345	49,345	48,240	48,237	48,237	48,237
R^2	0.011	0.062	0.264	0.279	0.287	0.292	0.297	0.486

begin by regressing IND1 on CLUSTER_RATIO without other controls or fixed effects. The coefficient estimate in column 1 is −0.212, indicating that IND1 is strongly and negatively related to CLUSTER_RATIO. That is, a higher level of geographical concentration of similar industry firms is associated with faster incorporation of industry-wide information into stock prices. In column 2, we next account for industry and year fixed effects. Although this mildly attenuates the magnitude of the coefficient on CLUSTER_RATIO, our inferences remain unchanged.

In the remainder of the columns in Table 2, we progressively account for other firm characteristics that might affect stock price informativeness. The speed at which stock price adjusts to information is also likely to depend on the liquidity of a firm's shares or the amount of investor interest. We therefore check whether the univariate relation we find in columns 1 and 2 is driven purely by the fact that firms outside industry clusters are, on average, smaller, less liquid, and less visible, as indicated in Table 1.

First, we control for firm visibility using three variables: LOG_SIZE, measured as the natural logarithm of the firm's market capitalization; LOG_NUM_EST, the number of analysts following the firm, measured as the natural logarithm of 1 plus NUM_EST; and total institutional ownership, LOG_NUM_INST, measured as the natural logarithm of the average total number of institutions holding the stock over the calendar year. We add each of these variables to the regression in columns 3–5 of Table 2. As expected, the coefficients of all three variables are negative and statistically significant, confirming that price delay is negatively related to investor interest. However, the coefficient estimate on CLUSTER_RATIO remains negative and significant. That is, controlling for the amount of investor interest a firm enjoys, being located near other local peer firms positively affects how fast the firm's stock price incorporates common industry information.

Next, in column 6 of Table 2, we control for the effects of liquidity on price delay. We include three commonly used liquidity measures: LOG_VOL, measured as the natural logarithm of the yearly average of monthly dollar trading volume; LOG_TURNOVER, measured as the natural logarithm of the yearly average of the monthly number of shares traded scaled by shares outstanding; and Amihud's (2002) illiquidity measure, LOG_ILLIQ, measured as the natural logarithm of the average daily absolute return scaled by daily trading volume.[7] The coefficient estimate on CLUSTER_RATIO remains negative and significant and shows little attenuation even after controlling for the level of stock liquidity in our analyses.

Finally, we consider other omitted effects that operate at either the MSA or firm level. For example, Loughran and Schultz (2005) show that geographic location systematically affects stock liquidity. Therefore, it is possible that our finding may be related to an omitted MSA or firm characteristic. To address this concern, in columns 7 and 8 of Table 2, we reestimate our regression model of industry price delay with the full set of controls, along with industry, year, and, alternately, MSA or firm fixed effects. In both instances, the coefficient estimate on CLUSTER_RATIO remains essentially unchanged.

In summary, we show in Table 2 that the greater presence of other local peer firms around a firm is positively related to how quickly the firm's stock price reflects common industry information. We also check the robustness of our finding using the two alternative industry price-delay measures. Similar to our analyses using IND1, we test whether industry price-delay measures IND2 and IND3 are systematically related to CLUSTER_RATIO after controlling for other firm characteristics and fixed effects. We present the estimation results in Table 3.

The coefficient estimates on CLUSTER_RATIO in all six columns of Table 3 are negative and significant. Whether we control for firm characteristics or include industry, year, or MSA fixed effects, the negative relation between industry price delay and CLUSTER_RATIO remains robust. Consistent with Table 2, after controlling for all other firm characteristics, a 1-standard-deviation increase in CLUSTER_RATIO reduces the delay with which stock prices adjust to industry-wide information by roughly 1% to 3%.

[7]In untabulated results, we included each liquidity variable individually. Our results remain unchanged. For brevity, we include in Table 2 only those specifications that include all three liquidity proxies together at the same time.

https://doi.org/10.1017/S0022109018000261 Published online by Cambridge University Press

TABLE 3
Robustness: Industry Clusters and Price Delay

In Table 3, IND2 and IND3 are the industry-wide price-delay measures constructed as in equations (3) and (4) in the text, respectively. CLUSTER_RATIO is defined as the number of firms in a given industry, as defined by the 3-digit Standard Industrial Classification (SIC) code, and in the same MSA code, scaled by the total number of firms in the same industry. LOG_SIZE is the natural logarithm of firm size, measured as market capitalization. LOG_NUM_EST is the natural logarithm of 1 plus the amount of analyst coverage for a firm in a given year, measured as the aggregate number of year-ahead earnings estimates as found in the Institutional Brokers' Estimate System (IBES) Detail History file. LOG_NUM_INST is the natural logarithm of the average of the total number of institutions currently holding the stock over the calendar year. LOG_VOL is the natural logarithm of the yearly average of monthly dollar trading volume. LOG_TURNOVER is the natural logarithm of the yearly average of the monthly number of shares traded divided by the number of shares outstanding. LOG_ILLIQ is the natural logarithm of the average daily absolute return over daily trading volume, constructed as in Amihud (2002). Standard errors are clustered by industry and are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: IND2			Dependent Variable: IND3		
Variable	1	2	3	4	5	6
CLUSTER_RATIO	−1.049***	−0.764***	−0.294**	−1.659***	−1.123***	−0.650***
	(0.76400)	(0.1761)	(0.1269)	(1.1230)	(0.21440)	(0.1250)
LOG_SIZE			0.027			0.044***
			(0.0210)			(0.0142)
LOG_NUM_EST			−0.162***			−0.173***
			(0.0190)			(0.0131)
LOG_NUM_INST			−0.106***			−0.115***
			(0.015)			(0.013)
LOG_VOL			−0.094***			−0.134***
			(0.0297)			(0.0245)
LOG_TURNOVER			0.113***			0.128***
			(0.0272)			(0.0306)
LOG_ILLIQ			0.067***			0.067***
			(0.0129)			(0.0085)
Year fixed effects	No	Yes	Yes	No	Yes	Yes
Industry fixed effects	No	Yes	Yes	No	Yes	Yes
MSA fixed effects	No	Yes	Yes	No	Yes	Yes
No. of obs.	48,342	48,342	47,938	48,874	48,874	48,468
R^2	0.003	0.048	0.090	0.020	0.099	0.272

IV. Causal Evidence from Relocations

In the previous section, we showed that the stock prices of firms located near other local peers reflect industry-wide information more quickly. Despite a wide array of controls, it remains possible that our finding may be driven by an omitted firm-level variable that is correlated with both industry clustering and price delay. To allay these concerns, we examine a subset of firms from our sample that exogenously relocate from one MSA to another. Because we can study the same firm before and after the relocation, we can better attribute the effects on price delay to the change in the firm's location.

To conduct this experiment, we begin by identifying firms that relocate during our sample period. As detailed previously in Section II, we merge our existing sample with the Compact Disclosure database, which provides the firm's geographical location by zip code. We map the zip codes of firms' headquarters into MSAs and then identify all firms whose headquarters have moved from one MSA into another over this period. Because we are attempting to measure changes in industry-cluster concentration, we exclude firms that have moved locally from one city to another within the same MSA. This leaves us with a sample of 465 moves between the years from 1990 to 2006.

Next, we ensure that these relocations are not related to a change in the firm's business in a manner that might affect its information environment other than the geography channel. That is, we remove "endogenous" moves from the relocation sample. This is critical because such moves confound the variation in geographic location with variation in other firm characteristics. For example, consider a firm that moves to Silicon Valley because of a strategic decision to refocus its business on technology. Analysts and money managers may choose to gather more information about this firm following its relocation due to the change in the firm's business strategy rather than the change in its location. The location change would simply be the endogenous outcome of the firm's decision to change its business and would confound our experiment. We need to exclude these types of relocations from our analysis in order to ensure that we capture the direct effects of geography on information production and price efficiency.

Thus, we hand-collect media articles on the 465 relocations to identify the stated reason for each move and then filter out endogenous firm relocations. For example, we remove the relocation of AppliedMicro from San Diego to Sunnyvale in 2005 because it was part of AppliedMicro's acquisition of 3ware, which was located in Sunnyvale. In contrast, we keep relocations that appear exogenous, such as Verilink's move from San Jose to Huntsville in order to reduce operating costs; Fair Isaac's relocation from San Jose to Minneapolis in order to be closer to executives' homes in Minnesota; and the move of Trico Marine Services from New Orleans to Houston following Hurricane Katrina. After filtering endogenous moves, 194 firm relocations remain in our sample.

Although the number of exogenous relocations is admittedly small, we are sacrificing sample size in order to examine the causal effect of geography on market efficiency. That is, we investigate whether firms that move into more concentrated industry clusters experience a reduction in their measures of price delay. Specifically, we regress IND1, IND2, and IND3 on a post-move dummy variable that takes a value of 1 in the calendar year following the year of the relocation, and we include year fixed effects to address general time trends in our price delay measures.

Table 4 shows the estimation results. The first three columns show the effect on the stock price efficiency of relocating into a smaller industry cluster. The coefficient estimates on the post-move dummy are positive for each of the three

TABLE 4

Price Delay for Relocating Firms

In Table 4, IND1, IND2, and IND3 are the industry-wide price-delay measures constructed as in equations (2), (3), and (4) in the text, respectively. Firm relocations are classified into smaller and larger clusters as measured by CLUSTER_RATIO. POST_MOVE_DUMMY takes a value of 1 in the calendar year following the year of the relocation, and 0 otherwise. Standard errors are clustered by firm and are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable					
	Moves into Smaller Clusters			Moves into Larger Clusters		
Variable	IND1	IND2	IND3	IND1	IND2	IND3
POST_MOVE_DUMMY	0.024	0.001	0.008	−0.080*	−0.566**	−0.571**
	(0.05)	(0.54)	(0.53)	(0.04)	(0.27)	(0.27)
No. of obs.	71	71	71	117	117	117
Adj. R^2	0.213	0.200	0.199	0.225	0.245	0.243
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes

https://doi.org/10.1017/S0022109018000261 Published online by Cambridge University Press

price-delay measures, but they are not statistically significant. Although the direction of the effect on delay appears to be consistent with what we expect, the lack of statistical significance may be due to the small sample size for this type of relocation in our sample.

In contrast, the coefficient estimates in the next three columns show that price delay drops significantly when firms relocate into a larger industry cluster. This effect is robust across all three price-delay measures. Consistent with our earlier findings, our analysis of relocating firms suggests that, holding all firm characteristics constant, being located in an area with greater industry concentration has a significant impact on the speed at which industry-wide information is impounded into stock prices.

V. Information-Spillover Hypothesis and Information Production for Firms Inside Clusters

In the previous sections, we presented evidence that firms' prices reflect industry information more quickly when those firms are located near industry peers. We next turn to why this might be the case. Our conjecture is that a group of similar firms that are co-located lowers the average cost of information acquisition for these firms. Given that peer firms around the same location are more likely to have similar fundamentals, learning about a firm in an industry cluster is likely to help in valuing other similar firms within the same area. We expect this spillover effect to lower the average cost of acquiring information about firms located inside industry clusters relative to those located outside industry clusters.

If cheaper information collection explains our evidence of more informative prices for firms residing within industry clusters, we expect the actions of information producers to reflect this fact. We consider two groups of information producers that are critical in the price discovery process, financial analysts and institutional investors,[8] and test whether their coverage and portfolio choices reflect a lowered cost of acquiring information for firms inside industry clusters.

A. Analysts

Our first test examines the relationship between the amount of analyst coverage a firm receives and the amount of industry clustering around the firm. Specifically, if being in the vicinity of other industry peers makes it easier and less costly to process the information about a firm for an analyst, then firms that are located in areas with higher levels of industry clustering should be more likely to be covered by analysts, ceteris paribus.

[8]Prior literature has found that financial analysts and institutional investors act as information intermediaries by contributing to the price-discovery process (Brown, Hagerman, Griffin, and Zmijewski (1987), Yan and Zhang (2007)). Analysts assimilate information from several sources (e.g., firm management; conference calls; and macroeconomic, industry-level, and financial statement analyses) and then disclose such information via earnings forecasts and stock recommendations. Institutional investors, conversely, use the information obtained from analysts' reports, and they augment it with their own in-house analysis (Cheng, Liu, and Qian (2006)), along with any information they may obtain through private communications with insiders (Khan and Lu (2013)). Whether analysts and institutional investors convey information through the disclosure of their reports or via their trading behavior, their actions increase the magnitude of information impounded into the firm's stock price.

We test this relationship in two ways in Table 5, where we measure the amount of analyst coverage for a firm in a given year as the aggregate number of year-ahead earnings estimates (NUM_EST) available from the Institutional Brokers' Estimate System (IBES) Detail History file. We begin with a univariate analysis. Using the continuous CLUSTER_RATIO variable to measure the

TABLE 5

Analysis of Analyst Coverage and Industry Clusters

Table 5 measures analyst coverage as the aggregate number of unique year-end estimates found within the Institutional Brokers' Estimate System (IBES) Detail History file. NUM_EST is the number of analysts following the firm, and LOG_NUM_EST is the natural logarithm of 1 plus NUM_EST. In Panel A, cluster ratios of low, medium, and high represent firms that reside in the least-, middle-, and most-concentrated industry clusters, as measured by CLUSTER_RATIO rankings of the lowest, middle, and upper terciles, respectively. INSIDE_CLUSTER represents a firm that is headquartered in a Metropolitan Statistical Area (MSA) with at least 10 other firms within the same 3-digit Standard Industrial Classification (SIC) code. OUTSIDE_CLUSTER represents all remaining firms not labeled as INSIDE_CLUSTER. *t*-statistics between differences are calculated using Welch's *t*-test. In Panel B, CLUSTER_RATIO is defined as the number of firms within the same 3-digit SIC code and MSA, scaled by the total number of firms in the same industry. CLUSTER_DUMMY is a binary variable that takes a value of 1 when a firm is headquartered within an MSA with at least 10 other firms within the same 3-digit SIC code, and 0 otherwise. LOG_SIZE is the natural logarithm of market value of equity computed from Center for Research in Security Prices (CRSP) data as the share price times the number of shares outstanding on the fiscal year-end date. PRC is the share price of the firm at the beginning of the fiscal year. RET is calculated as the 6-month returns prior to the end of the fiscal year. LOG_ADV is the natural logarithm of the sum of a firm's annual advertising and selling, general, and administrative (SGA) expenses, scaled by annual sales. LOG_BE_ME is the natural logarithm of the ratio of book equity to the market value of equity. Standard errors, clustered by firm, are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

Panel A. Univariate Analysis

	Dependent Variable: NUM_EST		
Variable	Mean	Standard Deviation	*t*-Test
CLUSTER_RATIO: Low	5.502	5.632	
CLUSTER_RATIO: Medium	6.095	6.093	
CLUSTER_RATIO: High	7.366	7.570	
Medium – Low	0.593***		6.476
High – Medium	1.270***		11.870
OUTSIDE_CLUSTER	6.054	6.030	
INSIDE_CLUSTER	6.781	7.289	
Inside – Outside	0.727***		8.382

Panel B. Regression Analyses

	Dependent Variable: LOG_NUM_EST	
Variable	1	2
CLUSTER_RATIO	0.137**	
	(0.058)	
CLUSTER_DUMMY		0.059***
		(0.010)
LOG_SIZE	0.481***	0.481***
	(0.008)	(0.009)
PRC	−0.002***	−0.002***
	(0.000)	(0.000)
RET	−0.627***	−0.629***
	(0.136)	(0.137)
LOG_ADV	−0.006	−0.007
	(0.014)	(0.015)
LOG_BE_ME	0.097***	0.097***
	(0.010)	(0.009)
Year fixed effects	Yes	Yes
Industry fixed effects	Yes	Yes
MSA fixed effects	Yes	Yes
No. of obs.	21,244	21,244
R^2	0.7048	0.7050

https://doi.org/10.1017/S0022109018000261 Published online by Cambridge University Press

amount of industry clustering in each MSA, we rank firms in our sample into three tercile groups of low-, medium-, and high-clustering groups and compare the average amount of analyst coverage across these three groups. We also repeat this analysis using the binary variable CLUSTER_DUMMY, as previously detailed in Section II.

We report univariate results in Panel A of Table 5. Firms that are located in areas more highly populated by their industry peers, as measured by CLUSTER_RATIO, have an average earnings estimate of 7.36, compared to 6.09 and 5.50 for the medium- and low-cluster groups, respectively. Differences in NUM_EST across the three groups are statistically significant at the 1% level. We obtain a similar result when we compare the average number of analysts across firms located inside and outside clusters. Firms that are located within a cluster of other industry peers have higher analyst coverage than firms that are not. These univariate results in Panel A provide preliminary evidence that analysts are, on average, more likely to cover firms located near their industry peers.

In Panel B of Table 5, we next test this relationship using panel regressions where, along with CLUSTER_RATIO and CLUSTER_DUMMY, we also include variables to control for other factors that may affect the amount of analyst coverage a firm receives. First, to control for the firm's overall environment, and hence to proxy indirectly for the average cost of collecting information on the firm for an analyst, we include firm size, as measured by the natural log of the firm's market capitalization. Second, because analysts have been shown to have a dual role as trade generators for their respective investment banks (Irvine (2000)) and a proclivity toward high-momentum glamour stocks (Lee and Swaminathan (2000), Jegadeesh, Kim, Krische, and Lee (2004)), we also control for the firm's book-to-market ratio (LOG_BE_ME) and price (PRC) as proxies for glamour, as well as the firm's past 6-month returns (RET) as a proxy for momentum.

Finally, we control for the firm's advertising expenditures to account for the possibility that investors could have higher demands for information about firms located within industry clusters. If banks allocate analysts to firms with the most investor demand, then our inferences regarding the effect of industry clustering on analyst coverage could be confounded by this alternative explanation. Following Grullon, Kanatas, and Weston (2004), who show that advertising is correlated with investor demand but is uncorrelated with new information about a firm, we include advertising expenditure, measured as the natural log of the sum of advertising and SGA expenses, scaled by the firm's sales revenue for the given year, as a proxy for investor demand.[9]

The dependent variable in both columns of Panel B in Table 5 is the natural log of 1 plus the number of analyst estimates, as previously defined. In column 1, the coefficient estimate on CLUSTER_RATIO is positive and significant at the 5% level. We obtain similar results using CLUSTER_DUMMY, which is positive and significant at the 1% level, suggesting that firms located in MSAs with a critical mass of at least 10 other peer firms have greater analyst coverage than

[9]Firms are not required to report advertising expenses, and prior literature has indicated that many firms that do not report advertising expenses separately choose to aggregate these expenses with other expenses in SGA.

isolated firms. The magnitude of this coefficient estimate suggests that moving into a cluster will result in a 1% increase in analyst following. We also note that the coefficient estimate on LOG_ADV is not significant in either specification. This gives us comfort that it is unlikely that the documented association between industry clustering and analyst coverage is driven by an alternative demand explanation.

In summary, the evidence in Table 5 suggests that firms located inside industry clusters, on average, have higher levels of analyst following, even after controlling for other factors that determine analyst coverage. This is consistent with our information-spillover explanation that analysts will be more attracted to firms inside clusters due to their lower cost of information acquisition.

B. Institutional Holdings and Mutual Fund Managers

We next examine the effect of industry clustering on the portfolio choices of institutional investors. Prior literature has documented that these investors acquire and trade on privately acquired information (Ke and Ramalingegowda (2005), Bushee and Goodman (2007)). As we previously argued, if similar peers located in the same area share common fundamentals, a fund manager who learns about firm A may also apply this information to firms B, C, and D in the same industry cluster. We therefore expect fund managers who learn of and choose to invest in firm A's stock to hold equity positions in firms B, C, and D as well.[10]

To test our prediction, we use the methodology of Bushee and Goodman (2007). We proxy a fund manager's choice to acquire private information about a firm by whether or not the manager has taken a large portfolio position within given a firm. Using 13F institutional holdings data from the Thomson Reuters database, which contains all money managers with more than $100 million under management, we construct an indicator variable BET that takes a value of 1 if the equity position invested in a firm falls into the top quintile of the institution's total holdings for the given year, and 0 otherwise.

We then test our prediction in two ways. First, we ask whether fund managers, on average, have a higher propensity to tilt their portfolios toward firms located in industry clusters. This would be consistent with fund managers having a preference for firms located in industry clusters if they expect to use their private information for other local peers. As a univariate test, we partition the MSAs in our sample into terciles by CLUSTER_RATIO and compare the average value of BET across firms in each cluster group.

Panel A of Table 6 reports the average BET value for firms located in low-, medium-, and high-cluster groups. Moving across terciles of CLUSTER_RATIO, BET increases monotonically from 18.3% to 21.4% to 24.3%. On average, money managers take larger positions in firms that reside in medium-cluster areas than in firms located in low-cluster areas, and they take even larger bets in firms located in highly concentrated clusters. These differences are statistically

[10]Because our data on institutional holdings come from 13F filings, our tests for learning only capture choices that result in long positions, as short positions are excluded from 13F disclosures. To the extent that information spillovers from industry clustering affect investment decisions on the long and the short side, the inability to test for information spillovers that result in short positions will weaken the power of our tests.

TABLE 6

Analysis of Institutional Holdings and Industry Clusters

In Table 6, the dependent variable BET equals 1 if the percentage of the institution's equity portfolio invested in a firm is in the top quintile of its total holdings for the given year, and 0 otherwise. In Panel A, cluster ratios of low, medium, and high represent firms that reside in the least-, middle-, and most-concentrated industry clusters, as measured by CLUSTER_RATIO rankings of the lowest, middle, and upper terciles, respectively. INSIDE_CLUSTER represents a firm that is headquartered in a Metropolitan Statistical Area (MSA) with at least 10 other firms with the same 3-digit Standard Industrial Classification (SIC) code. OUTSIDE_CLUSTER represents all remaining firms not considered as INSIDE_CLUSTER. t-statistics between differences are calculated using Welch's t-test. In Panel B, CLUSTER_RATIO is defined as the number of firms in a given industry, as defined by the 3-digit SIC code and the same MSA code, scaled by the total number of firms in the same industry. CLUSTER_DUMMY is a binary variable that takes a value of 1 when a firm is headquartered within an MSA with at least 10 other firms within the same 3-digit SIC code, and 0 otherwise. LOG_SIZE is the natural logarithm of firm size, measured as market capitalization. LOG_NUM_EST is the natural log of 1 plus the amount of analyst coverage for a firm in a given year, measured as the aggregate number of year-ahead earnings estimates as found in the Institutional Brokers' Estimate System (IBES) Detail History file. STDEV is the standard deviation of analysts' earnings forecasts for the upcoming year. LOG_VOL is the natural logarithm of the yearly average of monthly dollar trading volume. RET is calculated as the 6-month returns prior to the end of the fiscal year. LOG_ILLIQ is the natural logarithm of the average daily absolute return over daily trading volume, constructed as in Amihud (2002). Standard errors in logistic regressions, clustered by firm and year, are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively. The odds ratio for each independent variable is shown below the standard error.

Panel A. Univariate Analysis

	Dependent Variable: BET		
Variable	Mean	Standard Deviation	t-Statistic
CLUSTER_RATIO: Low	0.183	0.387	
CLUSTER_RATIO: Medium	0.214	0.410	
CLUSTER_RATIO: High	0.243	0.429	
Medium – Low	0.0303***		50.38
High – Medium	0.0292***		44.67
OUTSIDE_CLUSTER	0.2052	0.40380	
INSIDE_CLUSTER	0.2270	0.41891	
Inside – Outside	0.02196***		40.83

Panel B. Logistic Regression Analyses

	Dependent Variable: BET	
Variable	1	2
CLUSTER_RATIO	0.13380***	
	(0.02880)	
	1.14300	
CLUSTER_DUMMY		0.01980***
		(0.00637)
		1.02000
LOG_SIZE	0.54460***	0.54410***
	(0.00305)	(0.00305)
	1.72400	1.72300
LOG_NUM_EST	0.03210***	0.03190***
	(0.00487)	(0.00491)
	1.03300	1.03200
STDEV	−0.00741	−0.00742*
	(0.00399)	(0.00432)
	0.99300	0.99300
LOG_VOL	−0.09760***	−0.09660***
	(0.00561)	(0.00565)
	0.90700	0.90800
PRC	0.001020***	0.00101
	(0.00062)	(0.00062)
	1.00100	1.00100
RET	0.66800***	0.66830***
	(0.01450)	(0.01460)
	1.95000	1.95100
LOG_ILLIQ	−0.03550***	−0.03540***
	(0.00615)	(0.00615)
	0.96500	0.96500
Year fixed effects	Yes	Yes
Industry fixed effects	Yes	Yes
MSA fixed effects	Yes	Yes
No. of obs.	2,421,900	2,421,900

https://doi.org/10.1017/S0022109018000261 Published online by Cambridge University Press

significant at the 1% level. We repeat the same analysis by comparing the average BET value for firms that have at least 10 local industry peers (as identified by CLUSTER_DUMMY) to that of firms that do not. We obtain very similar results.

We next estimate a logistic regression with BET as the dependent variable. Along with CLUSTER_RATIO, we control for other variables that are likely to influence institutional investors' portfolio holding decisions. These include firm size, LOG_SIZE, as a control for the firm's information environment; stock price, PRC, volume, LOG_VOL, and Amihud's illiquidity measure LOG_ILLIQ to account for typical trade execution and microstructure concerns faced by larger-sized block trades; and finally, 6-month past returns, RET, to control for institutions that face window-dressing concerns (Lakonishok, Shleifer, Thaler, and Vishny (1991)), as well as those institutions that appear to have preferences for high-momentum stocks (Brunnermeier and Nagel (2004)).

In Panel B of Table 6, the coefficient on CLUSTER_RATIO in column 1 is positive and significant, suggesting that the concentration of other peer firms around a firm positively affects the probability that a manager places a large bet in a particular firm. This finding is consistent with our prediction that a fund manager is more likely to acquire information if his or her private information about one particular firm can be applied to more than one firm. In column 2, we replace our measure of industry clustering with CLUSTER_DUMMY and obtain similar results. The coefficient on CLUSTER_DUMMY suggests that moving inside of a cluster increases the amount of concentrated portfolio holdings for that firm by 9.64%.

We next ask whether placing large bets in a firm that is located inside an industry cluster affects the manager's equity holdings in other firms that are in the same industry cluster. Specifically, we test whether a fund manager's private information about a large firm located inside an industry cluster trickles to other smaller firms in the same area. This would be consistent with the evidence presented by Hou (2007), which shows an intra-industry lead–lag pattern resulting from the diffusion of information from large firms to small firms in the same industry.

To test this prediction, we begin by identifying firms that are in the top tercile of market capitalization among their peers within each MSA and label these as large firms; all remaining firms are then labeled as small firms. Next, we estimate the following panel regression:

$$\begin{aligned}(5)\qquad \text{NUM\_SMALL}_{i,j,t} \quad = \quad & \text{FIXED\_EFFECTS}_t \\ & + b_1\text{NUM\_LARGE\_BET\_INSIDE}_{i,j,t} \\ & + b_2\text{NUM\_LARGE\_BET\_OUTSIDE}_{i,-j,t} + e_{i,j,t},\end{aligned}$$

where $\text{NUM\_SMALL}_{i,j,t}$ is the total number of small firms that a manager holds in his or her portfolio in a given industry i in MSA j in year t; $\text{NUM\_LARGE\_BET\_INSIDE}_{i,j,t}$ is the total number of large firms in which the manager has taken a overweighted portfolio position in the same industry and MSA; and finally, $\text{NUM\_LARGE\_BET\_OUTSIDE}_{i,-j,t}$ is the total number of other large firms that the manager has overweighted in his or her portfolio in the same industry but not located in the same MSA. If a manager's private information about a large firm inside an industry cluster trickles to other smaller peer firms in the same area and affects the manager's portfolio decisions in these firms, we should expect a positive and significant coefficient estimate on NUM_LARGE_BET_INSIDE. In addition, to the extent that being geographically close to the large firm results in a more significant influence, we should expect the coefficient estimate on NUM_LARGE_BET_INSIDE to be greater than that on NUM_LARGE_BET_OUTSIDE.

We present our results in Table 7. First, in column 1, we estimate the regression in equation (5) using our full 13F sample.[11] Consistent with our prediction, we obtain a positive and significant coefficient estimate on NUM_LARGE_BET_INSIDE. This indicates that having an overweighted portfolio position in one more additional large firm located in an industry cluster is associated with having, in the same portfolio, equity holdings in an additional 1.193 small peer firms in the same industry cluster in a particular year. The positive and significant coefficient on NUM_LARGE_BET_OUTSIDE is consistent with evidence given by Hou (2007) on intra-industry information diffusion. We note, however, that its magnitude is much smaller and confirm that the difference in coefficient magnitudes between NUM_LARGE_BET_INSIDE and NUM_LARGE_BET_OUTSIDE is statistically significant at the 1% level. This suggests that the information spillovers from having a concentrated portfolio position in a large

TABLE 7

Institutional Holdings and Small Firms in Industry Clusters

In Table 7, the number of small cluster holdings is the aggregate number of firms held by the fund manager for a specific industry cluster and year, where the size for each firm is found in the lower two terciles when ranked within industry cluster and year. NUM_LARGE_BET_INSIDE is the aggregate number of firms held by the fund manager for a specific industry cluster and year, where the size for each firm is in the upper tercile when ranked within each industry cluster and year, and the firm has BET with value = 1, as previously described in Table 6. NUM_LARGE_BET_OUTSIDE is the aggregate number of firms held by the fund manager for a given year but outside of the specified industry cluster used to calculate NUM_LARGE_BET_INSIDE, where the ranked size for each firm is in the upper tercile when ranked within each industry cluster, and the firm has BET with value = 1. Standard errors are clustered by firm and are reported in parentheses. *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

	Dependent Variable: Number of Small Cluster Holdings		
Variable	1	2	3
NUM_LARGE_BET_INSIDE	1.1926***	0.8511***	0.9343***
	(0.3107)	(0.3174)	(0.3335)
NUM_LARGE_BET_OUTSIDE	0.1600***	0.1204***	0.1240***
	(0.0212)	(0.0079)	(0.0079)
Number of clusters	61	61	61
Industry fixed effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
MSA fixed effects	Yes	Yes	Yes
No. of obs.	1,676,352	79,638	71,089
R^2	0.2789	0.0920	0.1080
Data availability	1993–2006	2003–2006	2003–2006
Data source	13F data	Mutual funds data	Mutual funds data
Removal of firms where headquarters and fund have same MSA	No	No	Yes

[11]We defer discussion of columns 2 and 3 of Table 7 to the next section where we consider other alternative explanations.

firm in the same industry but outside the immediate geographical area is less pronounced than that of a concentrated portfolio position in a large firm in the same industry cluster. Overall, it appears that a firm's geographic proximity to industry peers provides information spillovers that ultimately influence managers' portfolio choices.

One possible concern with this finding is that it could be potentially driven by an unbalanced distribution of the number of firms in a given industry across different MSAs. For example, if 80 of the 100 total firms in the biotech industry were located in the San Diego area, whereas the remaining 20 were scattered throughout the rest of the country, then a sampling of firms by the fund manager may result in a mechanical correlation between the coefficient estimate on NUM_LARGE_BET_INSIDE and the dependent variable in equation (5). We address this concern by rescaling both NUM_LARGE_BET_INSIDE and NUM_LARGE_BET_OUTSIDE. We divide NUM_LARGE_BET_INSIDE by the total number of firms in the given industry in the given MSA and, similarly, divide NUM_LARGE_BET_OUTSIDE by the total number of firms in the industry but located outside the given MSA. This ensures that a random sampling of firms held by a fund manager should not generate a mechanical difference in the coefficient estimates. Our estimation results (untabulated) from this alternate specification remain similar to those in column 1 of Table 7, with no significant changes in our statistical or economic inferences. Consistent with our information-spillover explanation, institutional investors appear to make choices that suggest that they gravitate toward firms located within concentrated industry clusters and exploit the information commonalities that exist among these firms.

VI. Alternative Hypotheses and Robustness

A. Home Bias

In Section V.B, we present evidence that institutional investors have a higher propensity to hold stocks of firms located inside industry clusters. Because institutional investors are more likely to hold local stocks (Coval and Moskowitz (1999)), we next assess whether our results may be driven by this effect. To exclude this possibility, we first limit our sample to only those mutual funds that are *not local* to the industry clusters in which they invest. Specifically, using mutual fund summary data from CRSP, we match the MSAs of the locations of funds' headquarters with the MSAs represented in their portfolio holdings and delete all observations where the MSA of the holdings and that of the fund headquarters are identical. We then define NUM_SMALL, NUM_LARGE_BET_INSIDE, and NUM_LARGE_BET_OUTSIDE using the same classification as before and reestimate the regression in equation (5) using this sample.

We present these results in columns 2 and 3 of Table 7. For comparison, we first report in column 2 the estimation results we obtain with the full sample of mutual funds, and in column 3, we present our results with the home-bias-free sample. Neither column changes our inferences; the coefficient estimate on NUM_LARGE_BET_INSIDE remains positive and significant in both specifications. When we limit our sample to the entire universe of mutual funds in column 2, we find that a 1-unit increase in NUM_LARGE_BET_INSIDE results in holdings of 0.85 additional small firms within the same industry cluster in a given year. Deleting local funds, which correspond to roughly 11% of the mutual sample, our coefficient estimate remains positive and significant, 0.93. Similar to our analysis of the 13F institutional holdings data, we statistically test the difference between coefficient estimates on NUM_LARGE_BET_INSIDE and NUM_LARGE_BET_OUTSIDE and find it to be significant at the 1% level. These results provide comfort that our findings are unlikely to be caused by the home-bias effect.

B. SIC3-MSA Clusters as Finer Industry Classifications

Another possible explanation for our findings may be that industry clusters, that is, SIC3-MSA classifications, may be sorting firms into a finer industry classification than the standard 3-digit SIC classification (SIC3). For example, SIC3 code 333 refers to firms that specialize in nonferrous metals like copper and aluminum. If sorting firms into industry clusters also sorts them into finer industry groups, such as copper with the 4-digit SIC (SIC4) code 3331 or aluminum with the 4-digit code 3334, we might expect firms with the same SIC3-industry cluster to be more likely to have the same 4-digit SIC code relative to firms within the same 3-digit industry but outside the MSA location.

We test this hypothesis directly by examining the 4-digit SIC code of every firm located inside an industry cluster. Specifically, we check whether the probability of randomly drawing another firm with the same SIC4 code is higher for firms in the same industry cluster than it is for firms located outside. Table 8 tabulates these results by SIC3 code and shows that the probability of randomly drawing another firm with the same SIC4 code within the same industry cluster versus outside the cluster is statistically equivalent (t-statistic = 0.337).[12] We view these results as evidence that our findings are unlikely to be driven by the fact that our SIC3-MSA industry clusters are simply improved proxies for industry classifications relative to SIC3 codes.

C. Expertise/Effort Story

We also consider whether other potential theories based on expertise and effort could provide an alternative to the information-spillover conjectures. Van Nieuwerburgh and Veldkamp (2010), for example, develop a model where investors who focus their information-acquisition efforts on a smaller group of assets gain more precise knowledge about the assets' future payoffs. In our setting, this would suggest the possibility that fund managers prefer to acquire information about firms inside the same industry cluster because they can gain greater expertise by concentrating their learning efforts within a given industry and geographical region. Alternatively, managers might choose to learn about firms located in industry clusters because doing so might require less effort, for example, taking one flight to interview multiple managers in the same industry cluster.

[12]To ensure that these results are not driven by the mechanics of the SIC classification scheme, we rerun these analyses using 6-digit industry and 8-digit sub-industry Global Industry Classification Standard (GICS) codes, which are classified primarily on the basis of principal business activity (Bhojraj, Lee, and Oler (2003)). The results of our analyses (untabulated) with the GICS sample yield the same conclusions as our SIC3-MSA analyses and are available from the authors.

TABLE 8
Tests of Industry Clusters as Finer Industry Codes

For each firm recorded in Table 8, the probability of randomly drawing a firm within the same Standard Industrial Classification (SIC) 4-digit (SIC4) code cluster but outside of the 3-digit SIC (SIC3) code and Metropolitan Statistical Area (MSA) cluster (SIC3-MSA cluster) is calculated and then subtracted from the probability of drawing a firm with the same SIC4 code within the SIC3-MSA cluster. Results are aggregated across SIC3 code and across the entire sample (shown in bold). *, **, and *** indicate significance at the 10%, 5%, and 1% levels, respectively.

SIC3 Code	Increased Probability of Same SIC4 within Cluster	Standard Error	*t*-Statistic	SIC3 Code	Increased Probability of Same SIC4 within Cluster	Standard Error	*t*-Statistic
131	−0.00142	0.00527	−0.2694	381	−0.05090	0.04810	−1.0582
138	−0.01980	0.00859	−2.3050**	382	0.00611	0.01000	0.6110
201	−0.13020	0.11440	−1.1381	384	−0.04100	0.00513	−7.9922***
208	0.07000	0.05160	1.3566	394	−0.12900	0.05070	−2.5444***
209	0.06720	0.05500	1.2218	399	−0.06890	0.03830	−1.7990*
251	−0.12820	0.13710	−0.9351	504	−0.11530	0.01920	−6.0052***
271	−0.00322	0.07890	−0.0408	506	0.04620	0.02690	1.7175*
281	0.14560	0.05760	2.5278***	508	−0.01040	0.03770	−0.2759
282	0.05960	0.09470	0.6294	509	0.01230	0.03440	0.3576
283	0.01240	0.00333	3.7237***	512	0.16520	0.04320	3.8241***
284	0.07750	0.02640	2.9356***	513	0.09580	0.02640	3.6288***
289	0.03860	0.04910	0.7862	514	−0.07470	0.06430	−1.1617
291	0.04380	0.04100	1.0683	541	−0.03050	0.07310	−0.4172
314	0.09940	0.08250	1.2048	581	0.03870	0.01460	2.6507***
331	−0.03240	0.03950	−0.8203	596	−0.01780	0.02640	−0.6742
344	0.05080	0.07440	0.6828	599	−0.04270	0.03080	−1.3864
349	0.05820	0.05050	1.1525	701	0.06330	0.02650	2.3887***
353	0.20230	0.02620	7.7214***	731	0.02980	0.02380	1.2521
355	0.04950	0.02340	2.1154**	735	−0.12090	0.05140	−2.3521***
356	−0.02550	0.02820	−0.9043	736	−0.01480	0.02380	−0.6218
357	0.05190	0.00488	10.6352***	737	0.01640	0.00158	10.3797***
358	0.00811	0.05770	0.1406	738	0.00613	0.00915	0.6699
362	−0.08440	0.05410	−1.5601	799	−0.03800	0.02380	−1.5966
364	0.02330	0.05930	0.3929	806	0.08630	0.04940	1.7470*
365	0.02560	0.03870	0.6615	807	−0.02190	0.05770	−0.3795
366	−0.00796	0.00552	−1.4420	808	0.14410	0.03760	3.8324***
367	0.08990	0.00377	23.8462***	809	0.10550	0.01680	6.2798***
369	0.05620	0.03530	1.5921	873	−0.03180	0.01120	−2.8393***
371	−0.03660	0.02220	−1.6486*	874	0.04740	0.02230	2.1256**
372	−0.04750	0.04210	−1.1283				
				Mean	**0.01300**	**0.03860**	**0.3372**

In either of these alternative explanations, our results on fund managers' portfolio holding decisions in Table 7 would be strongest for those managers who hold assets within a single industry cluster and weaker for the other managers who have holdings in multiple clusters.

To test for these alternative mechanisms, we first create a dummy variable M_CLUSTER that takes a value of 1 if a manager has holdings in multiple industry clusters in a given quarter, and 0 otherwise. We then modify the regression specification in equation (5) to allow for an interaction between managers who hold stocks of firms in multiple industry clusters and the number of concentrated holdings that they have in large firms within those clusters. Specifically, we estimate the following equation:

$$
\begin{aligned}
(6)\qquad \text{NUM\_SMALL}_{i,j,t} \quad = \quad & \text{FIXED\_EFFECTS}_t \\
& + b_1\text{NUM\_LARGE\_BET\_INSIDE}_{i,j,t} + b_2\text{M\_CLUSTER} \\
& + b_3\text{NUM\_LARGE\_BET\_INSIDE}_{i,j,t} \times \text{M\_CLUSTER} \\
& + b_4\text{NUM\_LARGE\_BET\_OUTSIDE}_{i,-j,t} + e_{i,j,t}.
\end{aligned}
$$

If these alternative effort and expert theories are what primarily drives managers' choices to acquire information about firms inside industry clusters, we should expect a negative coefficient estimate on the interaction term NUM_LARGE_BET_INSIDE × M_CLUSTER because both the effort and expertise explanations suggest that the effect should be strongest for managers who concentrate all of their holdings within a single industry cluster. Conversely, if information spillover is the primary catalyst that motivates managers' decisions to hold stocks of firms in industry clusters, we should expect the effect to be strongest for managers who have holdings in multiple clusters because it would be more difficult to acquire expertise over multiple clusters. That is, under our information-spillover hypothesis, we should expect a positive coefficient estimate on the interaction term.

Using the same 13F sample as used in column 1 of Table 7, we find (in untabulated results) a positive and significant coefficient estimate on the interaction term ($b_3 = 0.65$, $p = 0.03$) and a positive but insignificant coefficient estimate for b_1 ($b_1 = 0.44$, $p = 0.34$). Taken together, these results appear to be most consistent with our information-spillover hypothesis and not supportive of other alternative explanations related to effort or expertise.

D. Changes in Information Production for Exogenous Movers

Similar to Section IV, we analyze the hand-collected set of exogenous movers to see whether firms that relocate into larger (smaller) industry clusters have higher (lower) levels of information production due to changes in the cost of information acquisition. Once again, we classify the set of movers into two groups: i) firm moves into a larger industry cluster, as measured by CLUSTER_RATIO, and ii) firm moves into a smaller industry cluster. We then examine the total change and percentage change in the information-production variables we considered earlier as a result of the relocation: ΔBET and ΔNUM_EST computed in the year of the firm's relocation as the level of BET and NUM_EST at the end of the calendar year, less the level of BET and NUM_EST at the start of the calendar year. We calculate PCTΔBET as ΔBET scaled by BET at the start of the calendar year, and we calculate PCTΔNUM_EST as ΔNUM_EST scaled by NUM_EST at the start of the calendar year.

In Table 9, we tabulate these results. For analyst estimates, we are unable to obtain statistical significance for either group of firm movers, likely due to the diminished sample size. Despite this, it is perhaps worth noting that ΔNUM_EST and PCTΔNUM_EST increase for firms that relocate into more concentrated industry clusters. For institutional holdings, we find that money managers are more likely to hold concentrated positions in firms that relocate into more concentrated industry clusters, with results being statistically significant for ΔBET and PCTΔBET at the 5% and 1% levels, respectively. The mean increase in the number of concentrated positions (ΔBET) is 7.235, a 77.3% increase as calculated by PCTΔBET. We do not find significant changes in ΔBET or PCTΔBET for firms that relocate into smaller clusters.

TABLE 9

Information Production for Relocating Firms

In Table 9, ΔNUM_EST and ΔBET reflect differences between the number of analyst estimates and the number of concentrated holdings at the beginning and the end of the calendar year in which the firm relocated. PCTΔNUM_EST and PCTΔBET are calculated as ΔNUM_EST scaled by NUM_EST at the beginning of the calendar year and ΔBET scaled by BET in the beginning of the calendar year, respectively. Firm relocations are classified into smaller and larger clusters as measured by CLUSTER_RATIO. Standard errors are reported in parentheses. *, **, and *** indicate significance of 2-tailed *t*-tests at the 10%, 5%, and 1% levels, respectively.

Panel A. Analyst Estimates

	Dependent Variable			
	Moves into Smaller Clusters		Moves into Larger Clusters	
Variable	ΔNUM_EST	PCTΔNUM_EST	ΔNUM_EST	PCTΔNUM_EST
Pre-post change	0.36 (0.39)	6.56% (0.065)	0.422 (0.34)	23.8% (0.14)
No. of obs.	20	20	51	51

Panel B. Institutional Holdings

	Dependent Variable			
	Moves into Smaller Clusters		Moves into Larger Clusters	
Variable	ΔBET	PCTΔBET	ΔBET	PCTΔBET
Pre-post change	2.105 (3.14)	2.33% (0.21)	7.235*** (2.95)	77.3%** (0.26)
No. of obs.	19	15	34	28

E. Alternative Measures of Industry Clustering

Finally, we examine whether our main results are robust to different measures of industry clustering and create alternative industry-cluster definitions based on 2-digit SIC classifications. We redefine industry clustering at the 2-digit SIC level in two different ways. First, we define a variant of CLUSTER_RATIO, CLUSTER_RATIO2, as the number of firms in a given industry, as defined by the 2-digit SIC code, and in the same MSA code, scaled by the total number of firms within the same industry. CLUSTER_RATIO2 is a coarser measure of industry clustering; it has a mean and median of 0.07 and 0.04, respectively, and ranges from 0.001 to 0.38 in our sample.

We also construct a market-value-based measure of industry clustering, VALUE_CLUSTER_DUMMY, to identify firms located inside industry clusters. VALUE_CLUSTER_DUMMY takes a value of 1 if a firm's MSA comprises 10% or more of the total market value of the firm's industry, as defined by 2-digit SIC code. Finally, we reestimate the regression in equation (1) using 2-digit SIC industry portfolios to construct price-delay measures, IND1, IND2, and IND3, with respect to industry-wide information.

To test whether our main result extends to different industry-clustering definitions using CLUSTER_RATIO2 and VALUE_CLUSTER_DUMMY, we reestimate specifications analogous to those in Table 2 with these alternative price-delay measures. We regress IND1, IND2, and IND3 solely on the industry-clustering measure, CLUSTER_RATIO2 and VALUE_CLUSTER_DUMMY, respectively, first without, and then with, industry, year, and MSA fixed effects. In 11 out of 12 specifications considered, the coefficient on the industry-clustering measure CLUSTER_RATIO2 or VALUE_CLUSTER_DUMMY is negative and significant at the 1% level (untabulated). In additional specifications, we add a battery of proxies for other firm characteristics and confirm that our main result holds at the 2-digit SIC level as well: Geographical industry clustering significantly improves the dissemination of information into stock prices.

VII. Conclusion

Much of the geography-based asset-pricing literature has focused on the relationship *between investors and the firms they invest in*. Locality reduces the cost of information for local investors, and hence, local investors, both retail and institutional, tilt their portfolios toward local stocks. In this article, we focus on the relationship between geographic clustering and the informativeness of market prices. We find that firms co-located in concentrated industry clusters have prices that incorporate industry information more quickly. To establish causality, we hand-collect a special set of "exogenous" firm relocations and find that firms that relocate into more concentrated industry clusters have lower levels of price delay when compared to their pre-move levels.

To explain these results, we propose an information-spillover hypothesis: Correlated information about firms inside industry clusters reduces the cost of acquiring information about these firms. We find support for our hypothesis in analysts' and fund managers' coverage and portfolio choices. We show that analysts are more likely to cover firms located within industry clusters; similarly, fund managers are more likely to hold stocks of firms inside industry clusters. We also find that fund managers are also more likely to hold other within-cluster firms when they have a large position in a firm that is already in the same cluster. Our evidence of the actions of these intermediaries is most consistent with our information-spillover hypothesis; information producers choose to gather information about groups, such as industry clusters, where they can use information about one firm to value another, thereby lowering their average costs of information acquisition.

References

Almazan, A.; A. Motta; S. Titman; and V. Uysal. "Financial Structure, Acquisition Opportunities, and Firm Locations." *Journal of Finance*, 65 (2010), 529–563.

Amihud, Y. "Illiquidity and Stock Returns: Cross-Section and Time-Series Effects." *Journal of Financial Markets*, 5 (2002), 31–36.

Audrestsch, D. B., and M. P. Feldman. "Knowledge Spillovers and the Geography of Information." In *Handbook of Regional and Urban Economics*, Vol. IV, J. V. Henderson and J.-F. Thiesse, eds. North-Holland (2004), 2713–2739.

Bhojraj, S.; C. M. C. Lee; and D. Oler. "What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research." *Journal of Accounting Research*, 41 (2003), 745–774.

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

Brown, L.; R. Hagerman; P. Griffin; and M. Zmijewski. "Security Analysts Superiority Relative to Univariate Time-Series Models in Forecasting Quarterly Earnings." *Journal of Accounting and Economics*, 9 (1987), 61–87.

Brunnermeier, M., and S. Nagel. "Hedge Funds and the Technology Bubble." *Journal of Finance*, 59 (2004), 2013–2040.

Bushee, B., and T. Goodman. "Which Institutional Investors Trade Based on Private Information about Earnings and Returns?" *Journal of Accounting Research*, 45 (2007), 1–31.

Cheng, Y.; M. Liu; and J. Qian. "Buy-Side Analysts, Sell-Side Analysts, and Investment Decisions of Money Managers." *Journal of Financial and Quantitative Analysis*, 41 (2006), 51–83.

Christ, J. P. "New Economic Geography Reloaded: Localized Knowledge Spillovers and the Geography of Innovation." FZID Discussion Papers 01-2009, University of Hohenheim, Center for Research on Innovation and Services (2009).

Coval, J., and T. Moskowitz. "Home Bias at Home: Local Equity Preference in Domestic Portfolios." *Journal of Finance*, 54 (1999), 2045–2073.

Coval, J., and T. Moskowitz. "The Geography of Investment: Informed Trading and Asset Prices." *Journal of Political Economy*, 109 (2001), 811–841.

Dougal, C.; C. Parsons; and S. Titman. "Urban Vibrancy and Corporate Growth." *Journal of Finance*, 70 (2015), 163–210.

Ellison, G.; E. Glaeser; and W. Kerr. "What Causes Industry Agglomeration? Evidence from Coagglomeration Patterns." *American Economic Review*, 100 (2010), 1195–1213.

Engelberg, J.; P. Gao; and R. Jagannathan. "An Anatomy of Pairs Trading: The Role of Idiosyncratic News, Common Information and Liquidity." Working Paper, Northwestern University (2008).

Engelberg, J.; P. Gao; and C. Parsons. "The Price of a CEO's Rolodex." *Review of Financial Studies*, 26 (2012), 79–114.

Grossman, S., and J. Stiglitz. "On the Impossibility of Informationally Efficient Markets." *American Economic Review*, 70 (1980), 393–408.

Grullon, G.; G. Kanatas; and J. Weston. "Advertising, Breadth of Ownership, and Liquidity." *Review of Financial Studies*, 17 (2004), 439–461.

Hameed, A.; R. Morck; J. Shen; and B. Yeung. "Information, Analysts, and Stock Return Comovement." *Review of Financial Studies*, 28 (2015), 3153–3187.

Hou, K. "Industry Information Diffusion and the Lead-Lag Effect in Stock Returns." *Review of Financial Studies*, 20 (2007), 1113–1138.

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

Irvine, P. "Do Analysts Generate Trade for Their Firms? Evidence from the Toronto Stock Exchange." *Journal of Accounting and Economics*, 30 (2000), 209–226.

Jacobs, J. *The Economy of Cities*. New York, NY: Random House (1969).

Jegadeesh, N.; J. Kim; S. Krische; and C. Lee. "Analyzing the Analysts: When Do Recommendations Add Value?" *Journal of Finance*, 59 (2004), 1083–1124.

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

Kedia, S., and S. Rajgopal. "Neighborhood Matters: The Impact of Location on Broad Based Stock Options Plans." *Journal of Financial Economics*, 92 (2009), 109–127.

Khan, M., and H. Lu. "Do Short Sellers Front-Run Insider Sales?" *Accounting Review*, 88 (2013), 1743–1768.

Lakonishok, J.; A. Shleifer; R. Thaler; and R. Vishny. "Window Dressing by Pension Fund Managers." *American Economic Review*, 81 (1991), 227–231.

Lee, C., and B. Swaminathan. "Price Momentum and Trading Volume." *Journal of Finance*, 55 (2000), 2017–2070.

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

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

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

Pirinsky, C., and Q. Wang. "Does Corporate Headquarters Location Matter for Stock Returns?" *Journal of Finance*, 61 (2006), 1991–2015.

Van Nieuwerburgh, S., and L. Veldkamp. "Information Acquisition and Under-Diversification." *Review of Economic Studies*, 77 (2010), 779–805.

Veldkamp, L. "Information Markets and the Comovement of Asset Prices." *Review of Economic Studies*, 73 (2006), 823–845.

Yan, X., and Z. Zhang. "Institutional Investors and Equity Returns: Are Short-Term Institutions Better Informed?" *Review of Financial Studies*, 22 (2007), 893–924.