Know Thy Neighbor:

Industry Clusters, Information Spillovers and Market Efficiency*

Joseph Engelberg

Arzu Ozoguz

Sean Wang‡

August 2013

Abstract: We find that firms in industry clusters – those in both the same industry and geographic area – have market prices that are more efficient than firms outside clusters. We argue this is the case because geography allows for information spillovers, reducing the marginal cost to information producers like analysts and fund managers. To support this view, we find firms inside (outside) clusters have fundamentals such as investment and earnings that have stronger (weaker) co-movement. We also find analysts are more likely to cover stocks inside industry clusters and that, when mutual fund managers have a large position in one stock in the industry cluster, they are more likely to hold other stocks in the same industry cluster. Finally, we examine a special set of exogenous firm relocations to make causal statements about the effect of geography on information production and market efficiency.

* We have benefited from discussions with Robert Bushman, Chris Malloy, Chris Parsons, and seminar participants at University of Southern California, Rutgers University, University of Washington, UT Dallas, Financial Intermediation Research Society Conference, Sixth Annual Asset Pricing Retreat (Amsterdam) and the American Finance Association Meeting.

‡ Contact: Joseph Engelberg, UCSD, Rady School of Management, jengelberg@ucsd.edu; Arzu Ozoguz, UT Dallas, Naveen Jindal School of Management, ozoguz@utdallas.edu; and Sean Wang, UNC, Kenan-Flagler Business School, sean_wang@kenan-flagler.unc.edu.

"After all, geographical proximity matters in transmitting knowledge because … intellectual breakthroughs must cross hallways and streets more easily than oceans and continents."

Handbook of Regional and Urban Economics (2004)

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 papers have provided evidence that locality reduces the cost of information (e.g. Coval and Moskowitz, 1999 and 2001, Loughran and Schultz, 2005). The idea is simple: a mutual fund manager located in San Diego will be able to collect information about each biotech firm in San Diego more easily 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.

This paper provides evidence of an *alternative* channel by which geographic proximity reduces the marginal cost of information: information commonalities among local firms. 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 information about the first biotech firm in San Diego to better understand the second biotech firm in San Diego at a lower cost. Building on this intuition[1], we ask whether geographic clustering at the industry level affects investors' choices in acquiring information and the efficiency of stock prices for firms located in industry clusters.

Of course, investors will only be motivated to learn about the marginal firm in the industry cluster if firms located in clusters behave in ways similar to their local peers. Fortunately, there is growing literature that provides evidence of such correlated behavior and the presence of local information networks. Kedia and Rajgopal (2008) find that locality explains patterns in option grants and argue that the social influence of neighboring firms drives their results. Dougal, Parsons, and Titman (2012) find that local firms have similar investment patterns even when they are in different industries. Pirinsky and Wang (2006) document strong

[1] See Veldkamp (2006) who formalizes this intuition. Specifically, she develops a rational model in which high fixed costs of producing information on individual firms cause investors to focus on signals that are common to many firms.

local co-movement in stock returns, showing that when a firm moves the location of its headquarters, returns exhibit stronger co-movement with firms in the new region. Engelberg, Gao, and Parsons (2012) provide evidence that CEOs are paid explicitly for their local network of connections to other executives and directors. The authors show that pay-per-connection is higher (lower) for local (remote) connections of the CEO which they argue is compensation for access to more valuable (less valuable) information networks. They also show that firms located inside (outside) industry clusters pay less (more) for each connection of their CEO, which they attribute to the fact that firms in industry clusters are endowed with local information networks. In urban agglomeration literature, many papers demonstrate 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 papers 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."

We also find empirical evidence of correlated behavior among firms located within industry clusters. Specifically, we examine the degree of co-movement of a firm's fundamentals with firms that are located in the same industry and geographic area. To the extent that managers within the same industry cluster make correlated decisions and are subject to the same local inputs to production, we predict a greater degree of co-movement in fundamentals amongst firms located within the same industry cluster.[2] Results confirm our predictions. Specifically, we find that fundamentals such as earnings and investment have stronger (weaker) co-movement inside (outside) industry clusters.

Having established that firms within industry clusters have correlated fundamentals, we next examine how this affects the actions of information intermediaries. We argue that the ability to use one piece of information to forecast the value of several firms will lower the average cost of learning, thereby attracting information intermediaries to firms that are headquartered within industry clusters. It is important to note, however, that this hypothesis about information intermedaries does not rely upon a specific channel to generate information commonalities among local firms. Any channel in the aforementioned papers (e.g. local labor

[2] How decisions are made when information is common to many firms is central to Scharfstein and Stein (1990), Welch (1992) and Rhodes-Korpf and Viswanathan (2004). These studies show how correlated decisions can result in herding and cascades and affect merger decisions.

markets, information sharing among local executives, etc.) is possible.[3] We provide robust evidence of correlated fundamentals amongst firms within industry clusters, and argue these correlations allow information intermediaries to use their knowledge about one firm in the cluster to better understand another.

In our analysis we focus on two groups of information intermediaries: financial analysts and institutional investors. We examine how the information choices of these two groups are affected by geographical clustering at the industry level. Consistent with our information spillover hypothesis, we find that analysts are more likely to cover a firm when that firm is located within an industry cluster.[4] Similarly, we find that institutional investors dedicate a higher percentage of their portfolio to firms that are headquartered within an industry cluster. Both of these results come from regression analyses that include industry fixed effects so that conclusions we reach are not driven by a firm's industry but instead where that firm is geographically located within an industry. We also aggregate the holdings of institutional investors' portfolios to examine their holdings *within* industry clusters. Our results show that fund managers who choose to gather information and 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 McKinlay (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 posit that lowered marginal information costs should also affect the market efficiency of stock prices for firms located within industry clusters. Using the price delay measure of Hou and Moskowitz (2005), we find that prices respond more (less) quickly to industry information inside (outside) industry clusters. Because we are concerned about possible omitted liquidity variables that jointly determine analyst coverage/institutional holdings and price delay, we include several different liquidity measures in our specifications. We also consider alternative formulations of the price-delay measure. Our main result – that

[3] Given extensive research within the urban economics literature regarding multiple mechanisms that may drive industry co agglomeration (see Ellison, Glaeser, and Kerr (2010) for a recent and in-depth discussion), it is extremely unlikely that there would exist a singular overarching mechanism that would drive the formation of clusters across the entire cross-section of industries. Regardless of the mechanisms that drive industry clustering, information producers still benefit from a lowered marginal cost of learning across firms within the cluster.
[4] This finding is also consistent with Hameed, Morck and Yeung (2008), who show that more analysts cover firms whose fundamentals are good predictors of many other firms' fundamentals.

prices for firms within industry clusters absorb information more rapidly, and are hence more informative -- is robust to all of our specifications.

Our most convincing tests consider a special set of firms that relocate to another geographic region. First, we use Compact Disclosure to identify all firm relocations between 1990 and 2006. From these 465 relocations we read the corresponding media accounts for the reason of the relocation and remove relocations that relate to a change in the firm's business or strategy which might affect 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. The remaining 194 "exogenous" relocations include 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 Trico Marine Services move from New Orleans to Houston following Hurricane Katrina. Such switches allow us to hold the firm constant but vary its presence in (or out) of an industry cluster.

Consistent with our prior findings, we find that following an exogenous corporate relocation: (1) firms demonstrate stronger co-movement with firms in the new industry cluster in investment fundamentals such as CAPX, R&D and SG&A, (2) firms that move from smaller to larger industry clusters have increases in analysts following, and an increased number of active institutional investors that possess concentrated holdings in their firm, and, (3) the stock prices of firms that move from smaller to larger industry clusters have improved levels of informational efficiency, as documented by lower levels of price delay.

We then consider the robustness of our main findings. First, we find that our results are not a rehash of Coval and Moskowitz (1999 and 2001), which find that local fund managers are more likely to hold local stocks. When we exclude local fund managers from the sample, we still find that a fund manager is more likely to hold a firm in an industry cluster if he or she already owns one. Second, when we look at the differences between SIC3 and SIC4 codes of firms in and out of industry clusters, we find no evidence that geography is simply a finer classification of industry. Thus, it is implausible that our results on correlated fundamentals, analyst coverage, and fund manager holdings are the result of a greater likelihood that clustered firms are members of a finer industry group. Finally, our main findings are robust to several different definitions of industry cluster and price delay.

Taken together, our results expand the role for geography in affecting investors' information choices and the cost of information which ultimately sets prices. With the well-documented home bias among institutional investors, locality reduces the cost of information *only for local funds*. In our paper, commonalities among local firms reduce the cost of information *for all funds*. This alternative explanation suggests that the relationship between geography and asset pricing may be more general than previously thought.

The paper is organized as follows. In Section II, we describe the data and the construction of variables, and in Section III, present some evidence of information spillovers via the co-movement of fundamentals among firms inside clusters. Section IV analyzes the impact of industry clusters on information intermediaries' actions to acquire information. In Section V, we compare the price delay of firms located inside and outside of industry clusters. Section VI discusses causality using results from a group of exogenous firm relocations. Section VII discusses alternative hypotheses, while Section VIII performs a set of robustness checks. Section IX concludes.

II. Data and Variable Construction

Our sample includes all NYSE-, AMEX- and NASDAQ-listed securities with share codes 10 or 11 that are contained in the intersection of Center for Research in Security Prices (CRSP) and COMPUSTAT databases from 1990 to 2007. We obtain return and pricing data from CRSP and merge it with accounting data from COMPUSTAT annual files, using the CRSP-LINK database produced by CRSP. To be included in our sample, we require for each firm to have (i) a non-missing SIC code in CRSP, and, (ii) state and county codes associated with the company headquarters from COMPUSTAT annual files. In addition, we exclude from our sample financial firms (SICs 6000-6999) and regulated utilities (SICs 4000-4999) as well as firms located in Hawaii and Puerto Rico. Finally, to minimize the effect of outliers, we winsorize all accounting variables at the 5% level.

Following Coval and Moskowitz (1999), Pirinsky and Wang (2006), and Almazan et al. (2010), we define each firm's locality as the geographical location of its headquarters. We use Metropolitan Statistical Areas (MSA) as defined by the 1990 United States Census Bureau to proxy for geographical location[5] and use state and county code classifications from COMPUSTAT annual files to merge our sample firms with the MSA codes.

[5] The general concept of a metropolitan statistical area, 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

While one concern with the COMPUSTAT location data is that COMPUSTAT only reports the *current* state and county of firms' headquarters,[6] in section 6, we specifically examine a subset of firms in our sample that relocate during our sample period. We overcome COMPUSTAT's shortcoming by merging our dataset with data from Compact Disclosure, which provides information on the zip code, city, and state of a 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 on MSA to another. Unfortunately, we have to limit our sample period to 1990-2006 for this analysis as this is the period for which we have Compact Disclosure data.

Throughout the paper, we focus on three-digit SIC classifications to define industry membership and clusters. This choice reflects a balance between the desire to minimize the possibility of grouping together firms in unrelated lines of business, while ensuring the viability of an 'industry cluster' definition. Although we present our results using three-digit SIC classifications throughout the paper, our findings remain qualitatively similar at the two-digit level.

To study the effect of geographic clustering on information spillover across firms and price efficiency, we first identify industry clusters. We measure the degree of industry clustering using a continuous variable *Cluster Ratio*, computed as the number of firms in a given industry, as defined by the three-digit SIC code, and in the same MSA code scaled by the total number of firms within the same industry. Higher levels of *Cluster Ratio* indicate more concentrated industry clustering. We also identify cluster-firms using a binary dummy variable, *Cluster Dummy*, which takes a value of one if a firm's MSA includes ten or more firms with the same three-digit SIC, and zero otherwise.

Our final sample includes 198 MSAs, 61 industries, 7,256 firms, and 52,697 firm-year observations. Table 1 provides descriptive statistics for the firms in our sample. Based on the cluster dummy variable, 34% of firms are located inside an industry cluster. Our industry cluster measure, *Cluster Ratio,* has an average value of 0.10, suggesting that, on average, 10 percent of firms within a 3-digit SIC code reside in the same MSA in our sample. This variable ranges from

---

Metropolitan Statistical Area definition which groups together a number of adjacent MSAs. A Consolidated Metropolitan Statistical Area has a population of one million or more and consists of separate components that are themselves considered Primary Metropolitan Statistical Areas (PMSA). For example, the San Francisco-Oakland-San Jose, CA CMSA consists of six 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.

[6] Pirinsky and Wang (2006) show that in the period 1992-1997, less than 2.4% of firms in COMPUSTAT changed their headquarter locations.

0.001 to 0.529 in our data (unreported). This number is 17 percent among MSAs with clusters, and drops to six percent when we condition on non-cluster areas.

In Table 2, we compare firms located inside clusters in our sample to those that are located outside clusters, as identified by the *Cluster Dummy*. *A*mong non-cluster firms, the average *Cluster Ratio* drops to 6 percent, while it increases to 15 percent among cluster firms. Table 2 suggests that cluster and non-cluster firms differ in several dimensions. We find that firms located inside clusters are slightly smaller in size, have significantly higher R&D, sales growth, cash balances and market-to-book ratios, and larger levels of intangible assets. They are also less profitable. These findings are largely consistent with Almazan, Motta, Titman and Uysal (2010).

III. Co-movement of Fundamentals

We begin our analysis by providing evidence on the extent of information commonalities among firms within industry clusters. We argue that if firms within industry clusters are exposed to common local shocks to production inputs and make correlated decisions, then ceteris paribus, we should observe increased co-movement in fundamentals among firms inside clusters relative to firms not inside industry clusters. In this section, we empirically examine the degree of this co-movement in our sample.

To measure co-movement, we relate the annual change in a firm's fundamentals $\Delta \underline{Fundamental}_{i,t}$, e.g. earnings, to the MSA-wide average annual change among same-industry firms located in the same MSA, denoted by $\Delta Fundamental_{MSA\text{-}IND}$, as well as to the industry-wide average change, $\Delta Fundamental_{IND}$. That is, we ask how much of the change in a firm's fundamentals can be explained by the change in other similar local firms' fundamentals, after controlling for industry-wide co-movement. We exclude the firm itself in computing both the local and industry-wide average change. Finally, to test whether a firm located inside an industry cluster has a higher degree of local co-movement, we include an interaction term between the average change in the local fundamental variable and the cluster dummy variable $Cluster_{i,t}$ that takes a value of one if the firm is located inside an industry cluster. Thus, our hypotheses regarding co-movement inside clusters are tested specifically with the coefficients on the interaction terms.

Table 3 shows the estimation results using various proxies to measure firm fundamentals. Specifically, we examine the co-movement, respectively, in firm profitability, measured by earnings and sales; in investment activity, measured by investment and R&D expenditures; and finally, in production inputs such as SGA Expenses and cash holdings. For each proxy, we estimate two specifications: one that leaves out the interaction term, and a second that allows the degree of local and industry-wide co-movement to differ for firms inside industry clusters. In all regressions, we also include year, industry and MSA fixed effects, and cluster standard errors by firm.

First, our results confirm that there are strong industry-wide commonalities in firm fundamentals. The coefficient estimates on $\Delta Fundamental_{IND}$ are all positive, ranging from 0.182 to 0.739, and they are statistically significant for all six of the proxies that we consider. Yet even after controlling for this industry-wide co-movement, there is evidence that firm fundamentals also co-move locally with the firm's local industry peers. There exists, on average, a positive, albeit weaker, relationship between a firm's fundamentals and those of its local peers in the same industry. The coefficient estimates on $\Delta Fundamental_{MSA\text{-}IND}$ are smaller in magnitude, but they are statistically significant for five of the six proxies that we consider. The degree of the local co-movement appears strongest for capital expenditures, R&D spending, and SGA expenses. This is consistent with the argument that the potential for influence among managers of same-industry firms increases significantly when firms are in close geographical proximity of each other.

Our focus in Table 3, however, is on the coefficient estimates on the interaction term $\Delta Fundamental_{MSA\text{-}IND}$ x *Cluster*. To the extent that firms within industry clusters are more likely to be exposed to common local shocks to production inputs and make correlated decisions, then ceteris paribus, we expect greater co-movement in fundamentals among firms inside clusters relative to firms not inside industry clusters.

We find that this is indeed the case. There is greater *local* co-movement among industry peers within an industry cluster, even after controlling for industry-wide effects. The coefficient estimates on the interaction term $\Delta Fundamental_{MSA\text{-}IND}$ x $Cluster_{it}$ are positive, ranging from 0.104 to 0.208, and they are strongly significant in all of the six columns for the different measures of firm fundamentals. That is, firms that reside within an industry cluster have stronger co-movement in their earnings, sales, investment and R&D expenditures, and even in their cash holdings with their *local* same-industry peers relative to firms that are located outside clusters. In fact, the coefficient estimates suggest the positive local co-movement in firm fundamentals can be almost entirely attributed to the firms located inside industry clusters in our sample. Incidentally, among the different proxies we consider to measure firm fundamentals, we find the strongest evidence of a greater degree of local co-movement inside industry clusters with SGA Expenses, which may indeed have the largest local component. This might be the case, for example, if firms inside industry clusters are exposed to similar local shocks to their production inputs and respond similarly.

Overall, in Table 3, using six different proxies of firm fundamentals, we show that firms inside industry clusters show significantly greater and more economically important co-movement with their local same-industry peers. These findings support our conjecture that firms which reside within industry clusters are likely to share greater information commonalities with their local peers.

While identifying the mechanism that underlies these commonalities is outside the scope of our study,[7] we conjecture these mechanisms will vary across different industries or geographical areas. We take comfort in the fact that the inclusion of industry, year and MSA fixed effects should address such concerns; therefore, our findings cannot be attributed to a specific MSA or industry effect. Finally, our results in this section suggest that learning about a single firm located in an industry cluster may inform about several others at the same time. We next ask whether this externality in learning affects information intermediaries' investment choices.

IV. Information Production for Firms Inside Clusters

Given the prior results in Section III that firms within clusters have information commonalities, we next ask whether the presence of positive information externalities within clusters affects information intermediaries' decisions to collect information. We posit that information intermediaries should have stronger incentives to collect information about firms that reside within industry clusters: the information gathered to value the first firm in the cluster can be used to value other firms within the same cluster, thereby lowering the average cost of information collection when calculated on a per firm basis.

[7] As we cite in the introduction, there are several studies in the literature that suggest information-sharing as one such mechanism.

Prior literature within capital markets research has found financial analysts and institutional investors do act as information intermediaries (Brown et al, 1987; Yan and Zhang, 2007). Analysts assist the price discovery process by assimilating information from several sources such as firm management, conference calls, macro-economic, industry-level and financial statement analyses and then disclosing such information via earnings forecasts and stock recommendations. Institutional investors, on the other hand, use the information obtained from analysts' reports and 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, 2011). 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. If the presence of information externalities among industry clusters indeed affects the incentives of analysts and institutional investors to focus on firms located within clusters, we should observe this effect, all else equal, in the coverage and holding choices of these agents.

IV.A. *Analysts*

Our investigation of analyst activity is centered on the amount of analyst coverage a firm receives. We begin with the assumption that information collection and processing is a costly endeavor (Grossman and Stiglitz, 1980). If this is the case, and information spillovers occur with firms that cluster in the same proximity thereby lowering the average cost of learning, then ceteris paribus, firms that reside in areas with more clustering will more likely be covered by analysts.

We measure the amount of analyst coverage for a firm in a given year by aggregating the number of year-ahead earnings estimates as found within the IBES Detail History file. Using the *Cluster Ratio* variable to measure the amount of industry clustering in each MSA, we rank the firms in our sample into terciles of high, medium, and low clustering groups, and conduct univariate analyses of the aggregate number of earnings estimates (NUMEST) across these groups. We also repeat the same analysis using the dummy variable *Cluster* that defines a firm to be inside a cluster if it is headquartered in a MSA with at least 10 other firms with the same 3-digit SIC code.

Results are reported in Table 4, Panel A and confirm our prior predictions. Firms in the upper tercile of clustering, as measured by *Cluster Ratio*, have an average of 7.36 earnings estimates, followed by an average of 6.09, and 5.50 for mid and low cluster groups, respectively. Differences in NUMEST across the three groups are statistically significant at the 1% level. When we compare the average number of analysts across firms inside and outside cluster, we obtain similar inferences, with firms inside clusters having an average of 6.78 earnings estimates per year, compared to an average of 6.05 available for firms classified as outside clusters.

Next, we estimate panel regressions where we regress the number of analysts following on each of the industry clustering measures, *Cluster Ratio* and *Cluster* respectively, and, at the same time, control for other factors affecting the expected level of analyst coverage. We include firm size (LogSize) and PIN, the probability of informed trading,[8] as proxies for the firm's information environment, hence indirectly proxying for the cost of information gathering for each analyst. Since analysts have been shown to have a dual-role as trade generators for their respective investment banks (Irvine, 2000), and proclivity towards high-momentum, glamour stocks (Lee and Swaminathan, 2000; Jegadeesh, Kim, Krische and Lee, 2004), we control for these effects by including in the regression the firm's book-to-market (LogBEME) ratio 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 entertain the possibility that investors could have higher demand for information about firms within industry clusters, and that banks allocate analysts to firms with the most investor demand. If true, then it is possible any inferences related to information sharing from the cluster concentration coefficients could be confounded by this demand story. Following Grullon et al (2004), that advertising is correlated with investor demand but uncorrelated with new information about a firm, we examine the relationship between a firm's advertising expenditures and analyst coverage. We calculate advertising (LogAdv) as the log of the sum of advertising and sales, general and administrative expenses, scaled by a firm's sales revenue for the given year.[9]

Table 4, Panel B shows the results of our multivariate regressions using the log of one plus the number of analyst estimates as the dependent variable, and confirms the inferences gathered from the previous univariate analyses. Column 1 shows the regression results using *Cluster Ratio*. The coefficient of *Cluster Ratio* is significant at the 5% level, and indicates that moving across the spectrum of *Cluster Ratio* will result in an increase of 1.4 analysts who issue earnings estimates for the firm for the upcoming year. Column 2 reports the results for the

[8] We thank Stephen Brown of the University of Maryland for providing us with PIN estimates.
[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 SG&A.

regression estimated with the *Cluster* dummy. Its coefficient is significant at the 1% level, and shows that firms located in MSAs with a critical mass of at least 10 other firms with the same 3-digit SIC code have a higher level of analyst following versus analyst following of isolated firms. In both regressions the coefficient on LogAdv is insignificant, making it unlikely that our inferences regarding *Cluster Ratio* and *Cluster Dummy* could be driven by an alternative demand story. In summary, consistent with our information spillover hypothesis, our analyses of analysts' coverage choices suggest that analysts may be attracted to firms inside clusters due to the lower per-firm average cost of information gathering and processing. After controlling for other factors related to the expected level of analyst following, firms whose headquarters are located inside (outside) industry clusters have higher (lower) levels of analyst following.

IV.B. Institutional Holdings & Mutual Fund Managers

Next, we examine the effect of industry clustering on a fund manager's portfolio holdings. We expect fund managers to be more inclined to learn about firms that are surrounded by other similar firms in the same industry; that is, located in concentrated industry clusters. For example, if a fund manager learns about Firm A, the correlated nature of this information can also help her learn about Firms B, C, and D within the same industry cluster. Given this assumption, we would expect fund managers who choose to learn about and hold Firm A to have holdings in Firms B, C, and D as well.

Using year-end 13F institutional holdings data obtained from the Thomson-Reuters database, we follow the methodology of Bushee and Goodman (2007): we proxy for a fund manager's choice to learn about and acquire private information about a firm by defining an indicator variable, BET, that proxies for a manager's large portfolio bets within a given firm. BET equals one if the percent of the institution's equity portfolio invested in a firm is in the top quintile of its total holdings for the given year and zero otherwise. We first test whether fund managers, on average, have a greater propensity to tilt their portfolios towards firms located in industry clusters by comparing the average value of BET across firms in high, medium and low cluster groups, as measured by *Cluster Ratio*. We also compare the average value of BET between firms inside and outside clusters, as defined by the *Cluster Dummy*.

Results are tabulated in Table 5 and are similar in spirit to our previous investigation of analyst coverage in section IV.A. Panel A documents that BET increases monotonically across *Cluster Ratio* ranks, with differences across cluster classifications being significant at the 1% confidence level. Economically, fund managers are 32% more likely to place big bets in firms that reside in areas with large industry clusters versus in isolated firms, increasing the proportion of their total portfolio holdings by 41% (untabulated) We obtain similar results using the dummy variable, *Cluster Dummy*.

In Panel B of Table 5, we show the results of logistic regressions using BET as the dependent variables. Firm size (LogSize) and PIN once again serve as initial controls for the informational environment, while stock price (PRC), volume (LogVOL) and Amihud's illiquidity proxy (LogILLIQ) are used to control for typical trade execution and microstructure concerns faced by larger-sized block trades. Six-month prior returns (RET) are used to control for institutions that face window dressing concerns (Lakonishok et al, 1991), as well as those institutions who appear to have strong proclivities toward high momentum stocks (Brunnermeier and Nagel, 2004). Results show that the coefficients on both of the clustering measures, *Cluster Ratio* or *Cluster Dummy* variable, remain positive and statistically significant in affecting the probability of a manager placing big bets in a particular firm. The odds that a firm is over-weighted by a given portfolio manager increases by 2.6% when a firm is within an industry cluster of at least 10 firms, and by 20.8% as the *Cluster Ratio* moves from 0 to 1.

Hou (2007) identifies an intra-industry lead-lag effect that results from diffusion of information from large firms to small firms in the same industry. Following this finding, it is interesting to ask whether placing large bets on a cluster firm affects a manager's holdings of other firms that also reside within the same industry cluster. Specifically, we first sort firms within each industry cluster by size, measured as market capitalization, and classify those that fall within the upper tercile as large, and all remaining firms as small firms. We then classify the large firms as being either *LARGE_FIRM_BET*, or *LARGE_FIRM_NOBET*, depending on whether or not the portfolio manager has overweighted the firm within her portfolio. Finally, we aggregate the holdings on a yearly basis and estimate a panel regression:

$$NUM\_SMALL_{i,j,t} = Fixed\ Effects_{\mathrm{t}} + \mathrm{b}_1\mathrm{NUM\_LG\_BET\_Inside}_{\mathrm{i,j,t}} + \mathrm{b}_2\mathrm{NUM\_LG\_BET\_Outside}_{\mathrm{i,-j,t}} + \mathrm{e}_{\mathrm{i,j.t}} \quad (1)$$

where, for a given manager-year with holdings in a particular industry cluster, $NUM\_SMALL_{i,j,t}$ is the total number of small firm holdings within a particular SIC3 code, MSA, and year. $NUM\_LG\_BET\_Inside_{i,j,t}$ is the aggregate number of large and overweighted firm holdings within a given SIC3 code, MSA and year, while $NUM\_LG\_BET\_Outside_{i,-j,t}$ is the total number of large and overweighted firm holdings in the same SIC3 code and year, but whose headquarters are located outside of the given MSA code. If, indeed, learning about a large cluster firm affects a manager's portfolio holdings in other firms in the same cluster, as predicted by our information spillover hypothesis, we would expect a positive and significant coefficient on b_1, and the coefficient on b_1 to be larger than that of b_2.

We present the regressions results in column 1 of Table 6[10]. As predicted, the coefficient on *NUM_LG_BET_Inside* is positive and significant. An overweight position in each additional firm within a given industry cluster implies holdings in 1.093 additional smaller firms within the same industry cluster for that particular year. In contrast, the coefficient on *NUM_LG_NOBET* is only 0.1138. A linear restriction test shows the coefficient on NUM_LG_BET_*Inside* to be significantly larger than NUM_LG_BET_*Outside* at the 1% significance level.

To address a possible concern that the differences in coefficients b_1 and b_2 may be driven by an unequal distribution of firms within a given SIC3 code across MSAs,[11] we rescale our independent variables from the original specification in (1). We divide *NUM_LG_BET_Inside* by the number of firms within the 3-digit SIC code inside the given MSA, and divide *NUM_LG_BET_Outside* by the number of firms within the 3-digit SIC code but outside of the given MSA. Under the modified specification, a random sampling of firms being held by a fund manager would result in equal coefficients of b_1 and b_2. Results from these untabulated analyses corroborate our primary specification, i.e. a larger coefficient on b_1 vs b_2, without any significant changes in either statistical or economic inferences.

Overall, analyses of institutional holdings appear to corroborate our findings with respect to analysts in support of the information spillover hypothesis. That is, sophisticated information intermediaries appear to act on the possibility of exploiting information commonalities in industry clusters by gravitating towards firms located within concentrated industry clusters.

V. Informational Efficiency Among Firms in Clusters

We next turn to the question of whether industry clustering has an important effect on informational efficiency of prices. The evidence we present so far suggests that geographic clustering at the industry level influences the decisions of information intermediaries by attracting them to firms within industry clusters. If that is the case, then consequently, we

[10] Columns 2 and 3 of Table 6 utilize mutual fund data to exclude the possibility that our results may be driven by the home bias effect (Coval and Moskowitz, 1999, 2001). We discuss these results in detail in Section VII.
[11] For example, assume that 100 firms were in the biotech industry, but 80 of the 100 were located in the San Diego area, while the other 20 were scattered throughout the US. Then, a sampling of firms by the fund manager might result in a mechanical correlation between the coefficient on *NUM_LG_BET_Inside,* and the dependent variable.

should expect a greater amount of industry-wide information to be also impounded into prices of firms inside clusters. To test this hypothesis, we empirically investigate whether firms located within industry clusters also have higher levels of stock price informativeness in our sample.

We proxy for the level of price informativeness by calculating the diffusion rates at which the firm's stock prices are able to incorporate industry-level information. Hou and Moskowitz (2005) create a price delay measure that examines the relation between individual stock returns and lagged market returns as a proxy for the rate at which market-wide information is incorporated into the price of an individual stock. Following their methodology, we construct several versions of the price delay measure to capture the speed of price adjustment to common industry information. If information intermediaries' focus on firms inside industry clusters makes their stock prices more informative, then firms located inside clusters should respond more quickly to common industry information than firms located outside of an industry cluster.

To compute the average delay of a firm's stock price with respect to common industry information, we regress each individual stock's weekly returns on the contemporaneous and four weekly lagged returns on the market and industry portfolios over the previous three years. Specifically, we estimate the following regression:

$$r_{i,t} = \alpha_i + +\beta_0 r_{M,t} + \sum_{n=1}^{4} \beta_n r_{M,t-n} + \delta_0 r_{IND,t} + \sum_{n=1}^{4} \delta_n r_{IND,t-n} + \epsilon_{i,t} \qquad (2)$$

where $r_{i,t}$ is the weekly return on stock i, and $r_{M,t}$ and $r_{IND,t}$ 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 (2)[12].

After estimating the coefficients in regression equation (2), we identify the delay in 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 than zero, but the lagged coefficients δ_n should not be different than zero. The first measure formalizes this intuition by measuring the fraction of the variation of the contemporaneous individual stock returns explained by lagged industry returns. That is, one minus the ratio of R^2 from regression (2) where $\delta_n, \forall n \in [1,4]$ are restricted to zero, to R^2 from regression (2) with no restrictions:

[12] See Engelberg, Gao and Jagannathan (2008) for a similar construction of an industry information diffusion measure.

$$IND1 = 1 - \frac{R^2_{\delta_n = 0, \forall n \in [1,4]}}{R^2} \quad (3)$$

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.[13]

Our objective is to understand whether being located within an industry cluster is associated with greater price informativeness, or equivalently, smaller price delay. To address this question, we begin by estimating a linear regression of price delay measures on our industry cluster measure *Cluster Ratio*. Table 7 presents the results of estimating a number of such specifications with various control variables. In all regressions, we cluster the standard errors at the industry level.

First, in column 1 of Table 7, we regress the price delay measure *IND1* solely on *Cluster Ratio*. The coefficient estimate in column 1 indicates the price delay measure *IND1* is strongly and negatively related to *Cluster Ratio*. That is, stock prices incorporate industry-wide information faster in an industry cluster as the number of firms in a cluster increases. Accounting for industry and year fixed effects in column 2 mildly attenuate the coefficient on *Cluster Ratio,* but leaves inferences unchanged.

In columns 3 through 7, we progressively add to the regression controls for firm characteristics that are expected to affect stock price informativeness. This is important since price delay can arise for various other reasons such as a lack of liquidity in a firm's shares, or a lack of investor interest. We therefore need to understand 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 seen in Table 2.

First, we try to control for visibility by including in columns 3 through 5, progressively, *firm size*, measured as the logarithm of the firm's market capitalization, the *number of analysts* following the firm, measured as the logarithm of one plus *NUMEST*, and total institutional ownership (*LOGNUMINST*), measured as the logarithm of the average total number of institutions holding the stock over the calendar year. As we expected, the coefficient estimates on all three variables are negative and statistically significant, confirming that price delay is negatively related to investor interest. However, even after controlling for the degree of investor attention, our main result – that firms in industry clusters respond faster to industry-wide information – remains unchanged. While the magnitude of the coefficient on *Cluster Ratio* is

[13] We describe the other two versions of the delay measures in section VI.B, where we discuss the robustness of our results.

reduced after controlling for investor interest, it remains strongly significant and negative. That is, being surrounded by a larger number of local same-industry peers appears to have a negative effect on price delay beyond that of the average effect of investor interest.

In column 6, we also add in liquidity measures to control for the effect of liquidity on price delay. We employ one of three different liquidity measures commonly used in the literature: the *average dollar trading volume*, measured as the logarithm of the yearly average of monthly dollar trading volume; *share turnover*, measured as the logarithm of the yearly average of the monthly number of shares traded divided by number of shares outstanding; and Amihud's (2002) *illiquidity* measure, measured as the logarithm of the average daily absolute return over daily trading volume[14]. The estimation results in column 6 show that the magnitude of the coefficient on *Cluster Ratio* shows very little attenuation when various liquidity controls are added.

Finally, one remaining possible concern may be that our main result is driven by some omitted fixed effects at the MSA level. We address this concern in column 7 by adding to the regression MSA fixed effects along with industry and year fixed effects as well as all of the controls for firm characteristics considered in columns 1 through 6. The coefficient on *Cluster Ratio* barely changes and our inferences remain the same as before, even after controlling for a batter of liquidity proxies and fixed effects.

Overall, Table 7 confirms that geographic clustering at the industry level facilitates the stock price adjustment to industry-wide information. This is consistent with the idea that the presence of information spillovers across firms within industry clusters also yields faster diffusion of industry-wide information, and hence, more informationally efficient stock prices.

VI. Causal Evidence from Relocations

In the previous section, we present evidence that suggests geographical industry clustering facilitates faster diffusion of industry-wide information into firms' stock prices. One concern may be that we have not adequately controlled for all potential firm characteristics that may influence stock price informativeness, and hence, results could be driven by omitted

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

variables. To address this concern, we examine a special subset of firms in our sample that relocate from one MSA into another.

The main problem with identifying firms that have relocated during our sample period is that COMPUSTAT only reports the current state and county of firms' headquarters. To overcome this issue, we merge our dataset with data from the Compact Disclosure database, which provides information on the zip code, city, and state of a firm's headquarters on an annual basis. Unfortunately, the Compact Disclosure data covers only a part of our sample period; we therefore limit this analysis to the period from 1990 to 2006. Using Compact Disclosure data, we first annually 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. Focusing on firms that change MSAs, we exclude firms that have moved locally from one city to another within the same MSA. Our final relocation sample consists of 465 migrations over the period 1990 to 2006.

From these 465 relocations we hand collect and read the corresponding media accounts for the reason of the relocation and remove relocations that relate to a change in the firm's business or strategy which might affect 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. The remaining 194 "exogenous" relocations include 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 Trico Marine Services move from New Orleans to Houston following Hurricane Katrina.

Removing "endogenous" moves is critical to this analysis because they confound variation in geography with variation in other firm characteristics. Suppose, for example, a firm moves to Silicon Valley because it wants to change its business to focus on technology. Analysts and institutions may gather more information about the firm because of the firm's change in business, not the change in location. The location change is simply the endogenous outcome of the firm's decision to change its business. These are precisely the firm relocations we exclude from our analysis. The remaining relocations which occur for exogenous reasons allow us to make causal statements about the effect of geography on information production and price efficiency.

We begin our analysis of relocations by examining the co-movement in fundamentals for this set of firm relocations. Specifically, if geography impacts a firm's corporate decisions, then we would expect the firm's fundamentals to have increased co-movement with the new location in the years following the firm's move. Using the same variables as those in section III, we examine the relation between changes in firm fundamentals (investment, profitability, SGA expenses, and cash) with the change in fundamentals of the firm's new industry cluster. In Table 8, Panel A, we examine these univariate correlations in the two years prior to the move, the move year, and the two years following the move. Overall, our results are consistent with prior literature (Dougal, Parsons, and Titman 2012) and show that a firm's investment is most strongly correlated with the new location in the years following the move. $\Delta CAPX$, $\Delta(CAPX + RD)$, and ΔSGA are all largest in the post-move years, with statistical significance at the 1% level.

In Panel B, we document results of multivariate regressions with the following specification:

$$\begin{aligned}\Delta Fundamental_{i,t} = {} & FixedEffects + \beta_1 \Delta Fundamental_{MSA,IND,OLD} + \beta_2 Postmove + \\ & \beta_3 \Delta Fundamental_{MSA,IND,OLD} Postmove + \beta_4 \Delta Fundamental_{MSA,IND,NEW} + \\ & \beta_5 \Delta Fundamental_{MSA,IND,NEW} Postmove + \beta_6 \Delta Fundamental_{IND} + \varepsilon_{i,t}\end{aligned} \tag{4}$$

Specifically, equation (4) regresses the change in the firm's fundamentals on changes in the fundamentals of both the pre- and post-move industry clusters, interacted with a dummy variable for the years following the firm's relocation, and the average change in the fundamentals of the entire industry. Given our expectations that the firm's fundamentals will co-vary more strongly with the new location in the post-move years, we expect β_5 to be positive and significant even after controlling for fixed effects and non-location based industry co-movement. Results are very similar to the univariate analyses in Panel A. The coefficient on β_5 is positive and significant for $\Delta CAPX$, $\Delta(CAPX + RD)$, and ΔRD, and positive but insignificant for ΔSGA, again suggesting that the firm's changes in investment are most strongly correlated with the new location after the firm has moved its headquarters. It is interesting to note that increased co-movement in profitability is not immediately seen in the post-move era, as is found in Table 3. This may be a result of the fact that changes in investments will lag changes in future profits, and that that the event window following the move is too narrow to capture the co-movement in profitability.

We follow our analyses of fundamental co-movement by examining changes in the activity of information producers—analysts and institutional investors. Following our results in section IV, if being located in a more concentrated industry cluster lowers the marginal cost of information acquisition through sharing greater information commonalities with other local peers, we expect to find increased information production from analysts and institutional investors following a firm's relocation into a more concentrated cluster. To test this hypothesis, we first classify the set of exogenous relocations we have 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 change in the three information production variables we considered earlier as a result of the relocation: ChgBET and ChgLogNUMEST are computed in the year of the firm's relocation as the level of BET and logNUMEST at the end of the calendar year, less the level of BET and logNUMEST at the start of the calendar year, while ChgBETPCT is calculated as ChgBET scaled by the level of BET at the calendar year.

Table 9, Panel A documents results showing increased activity for firms that move into larger industry clusters. On a univariate level, we find that firms moving into large clusters exhibit larger changes in ChgBETPCT relative to firms moving into small clusters, significant at the 5% level. Similarly, we find the increase in analyst following is higher for firms that move into large clusters versus small clusters, although the results are statistically insignificant due to our diminished sample size. Panel B shows results from regressing the change in information production on a relocation dummy which takes a value of 1 (-1) when a firm moves into a larger (smaller) industry cluster. After controlling for year fixed effects, the coefficient of the relocation dummy shows that a firm moving into a larger (smaller) industry cluster will have an increase (decrease) in BET of 4.80, a 42% increase in BETPCT, and 1.42 analysts who follow the stock. All three coefficients are positive and significant at the 5% level, and suggest that the information environment is improved for these firms.

Finally, we check whether firms that move into larger industry clusters also experience an improvement in their price efficiency. If, indeed, firms benefit from a higher level of attention from information producers when they are surrounded by a larger number of local peers in the same industry, we should also expect a greater amount of industry-wide information to be impounded into their stock prices after the relocation. That is, we should expect to see a lower level of price delay subsequent to the relocation.

We test this hypothesis by, again, comparing the pre- and post-move levels of the price delay measure for both types of relocations. Specifically, we regress price delay measure, *IND1*, on a post-move dummy variable that takes a value of one in the calendar year following the year of the relocation, controlling for year fixed effects. For robustness, as in Hou and Moskowitz

(2005), we also consider alternative definitions of price delay to make sure our results are not sensitive to how we measure price delay. Specifically, we construct *IND2* and *IND3,* which attempt to distinguish between shorter and longer lags and the precision of the estimates,

$$IND2 = \frac{\sum_{n=1}^{n=4} n\delta_n}{\delta_0 + \sum_{n=1}^{n=4} \delta_n} \tag{6}$$

and

$$IND3 = \frac{\sum_{n=1}^{n=4} n\left[\frac{\delta_n}{se(\delta_n)}\right]}{\frac{\delta_0}{se(\delta_0)} + \sum_{n=1}^{n=4}\left[\frac{\delta_n}{se(\delta_n)}\right]} \tag{7}$$

where $se(\cdot)$ denotes the standard error of the coefficient estimate.[15]

Table 10 shows the estimation results. We find that price delay drops significantly following the relocation for firms that move into larger industry clusters for all three measures of price delay. This is consistent with our earlier findings and suggests that, holding all other firm characteristics constant, being located in an area with greater industry concentration has a significant impact on how fast industry-wide information gets impounded into stock prices. It is interesting to note, however, that we find no significant change in average price delay when firms relocate into smaller clusters.

VII. Alternative Hypotheses

VII.A. Home Bias

In section IV, we find a tendency of institutional investors to hold stocks within an industry cluster. Because institutional investors like to hold local stocks (Coval and Moskowitz, 1999), our results may be driven by this home bias effect. To exclude this possibility, we perform again the test of Section IV only among mutual funds that are *not local* to the industry clusters they invest in. Specifically, we match the MSAs of the location of each mutual fund's headquarters with the MSAs of their given portfolio holdings, then delete all observations where the MSAs of the holdings and the firm headquarters are identical. We then create the same

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

classifications, using *NUM_SMALL*, *NUM_LG_BET_Inside*, and *NUM_LG_Outside* as in the previous analysis of the 13F firms. We report the results of the panel regression in Table 6. Column 2 shows results using the entire sample, while column 3 shows regression results after the deletion of "home-bias" firms. Inferences remain unchanged, with *NUM_LG_BET_Inside* being positive and significant in both specifications. For this smaller subset of mutual fund data, an additional unit of increase in *NUM_LG_BET_Inside* would result in holdings in 0.77 additional smaller firms within the same industry-cluster for a given year. After deleting the home-bias firms (roughly 11% of the mutual-fund sample), the coefficient is 0.85. Following our analyses on the 13F dataset, we run linear restriction tests to compare the differences in the coefficients of *Num_Large_BET_Inside* and *Num_Large_BET_Outside*. The results of our linear restriction tests illustrate the differences on these coefficients to be significant at the 1% level, and add further robustness to our information spillover hypothesis from section IV.

VII.B. SIC3-MSA Clusters as Finer Industry Classifications

One possible alternative explanation for our results is that industry clusters (SIC3-MSA classifications) might sort firms into a finer industry classification than the standard 3-digit SIC3 classification. For example, SIC3 code 333 refers to firms that specialize in non-ferrous metals like copper and aluminum. It may be that sorting firms into industry clusters also sorts them into finer industry groups like copper (4-digit SIC 3331) or aluminum (4-digit SIC 3334). If this were indeed the case, then we would expect that firms within the same SIC3-industry cluster would be more likely to have the same 4-digit SIC code relative to firms within the same SIC3 but outside of the MSA location.

We test this hypothesis directly by examining the SIC4 code of every firm within a given SIC3-MSA industry cluster, and calculating whether the probability of randomly drawing another firm with the same SIC4 code is higher for firms within the same industry cluster, versus outside of the industry cluster. Table 11 tabulates 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 of the given industry cluster is statistically equivalent (t-stat = 0.337).[16] We view these results as evidence that our results are unlikely to be driven by the fact

[16] 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, Oler, and Lee, 2003). Results of our analyses (untabulated) with the GICS sample yield the same conclusions as our SIC3-MSA analyses, and are available upon request.

that our SIC3-MSA industry clusters are simply improved proxies for industry classifications relative to SIC3 codes.

VII.C. Expertise/Effort Story

Finally, we consider alternatives based upon expertise and effort to our information spillover story. Regarding expertise, Van Nieuwerburgh and Veldkamp (2009) develop a model suggesting that investors who focus their information acquisition efforts on a smaller group of assets may be rewarded with more precise knowledge about their future payoffs. We consider the possibility that fund managers may prefer to acquire information about firms within the same industry cluster because they are able to gain expertise by concentrating learning efforts within a given industry and geographical region. On the other hand, managers might choose to learn about firms within a given industry cluster because they would be able to do so with less effort (i.e. they can take one flight to interview all of the managers within a given industry cluster). In either case, both alternative stories would suggest the results from our fund manager analyses in Table 6 would be strongest for managers that only hold assets within a single industry cluster, and weaker for managers that hold multiple clusters.

We test this alternative by creating a dummy variable that takes a value of 1 if a manager holds multiple industry clusters within a given quarter, MCLUSTER, and modifying the specification in equation (1) with an interaction between managers that hold firms in more than one industry cluster and the number of concentrated holdings they have in large firms within those clusters.

$$\begin{aligned} NUM\_SMALL_{i,j,t} &= Fixed\,Effects_t + b_1 NUM\_LG\_BET\_Inside_{i,j,t} + b_2 MCLUSTER_{i,t} \\ &+ b_3 NUM\_LG\_BET\_Inside_{i,j,t} MCLUSTER_{i,t} + b_4 NUM\_LG\_BET\_Outside_{i,j,t} + e_{i,j,t} \end{aligned} \quad (5)$$

If effort and expertise are the primary drivers behind managers' choices to learn about firms within industry clusters, results of the regression in equation (5) would be strongest for managers who concentrate all of their holdings within a single industry cluster, implying a negative coefficient for b_3. On the other hand, if information spillovers are the primary catalysts behind managers' decisions to hold firms within a given industry cluster, results would be strongest for managers that hold multiple clusters—as it would be more difficult to acquire expertise over multiple clusters. This story would imply a positive coefficient for b_3. Untabulated findings using the same 13F dataset as the prior analyses show b_3 to be positive and significant ($b_3 = 0.65$, $p = 0.03$), whereas b_1 is insignificant ($b_1 = 0.44$, $p = 0.34$). Taken together, results appear to be consistent with the information spillover hypothesis, while ruling out the possibility that our results are driven by alternative stories related to effort and expertise.

VIII. Robustness

VIII.A. Alternative Measures of Price Delay

Although we present our main price-delay results using *IND1* in section V, we obtain similar results in analyses using the alternative measures of price delay, *IND2* and *IND3*. We re-estimate the specifications considered in Table 7 with these delay measures.

In untabulated results from regressing *IND2* and *IND3* on our industry cluster measure *Cluster Ratio*, with industry, year, MSA fixed effects, and an array of firm characteristics, we find the coefficient on *Cluster Ratio* is negative and significant at the 1% level, confirming our finding in the previous section. We obtain very similar results when we use either IND2 or IND3. Regardless of specification, firms inside industry clusters experience less price delay than firms outside clusters.

VIII.B. Alternative Measures of Industry Clustering

In this section, we examine whether our results are robust to different measures of industry clustering and extend to cluster definitions based on the two-digit SIC classifications. Specifically, we measure the degree of industry clustering at the two-digit SIC level in two different ways. First, we define a variant of *Cluster Ratio*, *Cluster Ratio-2,* as the number of firms in a given industry, as defined this time by the *two-digit* SIC code, and in the same MSA code, scaled by the total number of firms within the same industry. *Cluster Ratio-2* 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 one if a firm's MSA comprises ten percent or more of the total market value of the firm's industry, as defined by two-digit SIC code. Finally, we re-estimate regression equation (2) using two-digit SIC industry portfolios to construct price delay measures, *IND1*, *IND2* and *IND3*, with respect to industry-wide information.

In order to test whether our main result extends to different industry clustering definitions using *Cluster Ratio-2* and *Value Cluster Dummy,* we re-estimate specifications analogous to those in Table 7 with these alternative price delay measures. We regress *IND1*, *IND2* and *IND3*, solely on the industry clustering measure, *Cluster Ratio-2* and *Value Cluster Dummy*, respectively, first without, and then with, industry, year and MSA fixed effects. In (untabulated) 11 out of 12 specifications considered, the coefficient on the industry clustering measure *Cluster Ratio-2* or *Value Cluster Dummy* is negative and significant at the 1% level. In additional specifications, we add a battery of proxies for other firm characteristics, and are able to confirm that our main result holds at the two-digit SIC level as well: geographical industry clustering significantly improves the dissemination of information into stock prices.

IX. 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 paper, we focus on the role of geography *among firms* and argue that correlated information reduces the marginal cost of information. Hence, information producers will tend to gather information about groups - like industry clusters – where they can use information about one firm to value another.

Our evidence comes in three forms: (1) we show that firms within an industry cluster have more correlated fundamentals (e.g. earnings and investment) than firms outside industry clusters, (2) we show that analysts and fund managers are more likely to cover/hold firms within industry clusters and that fund managers are more likely to hold other within-cluster firms when they have a large position in a clustered firm, and, (3) firms within industry clusters have slower price delay with respect to industry information. In our most precise tests we examine a special set of "exogenous" firm relocations so that we can make causal statements about the effect of geography on information production and efficiency.

Because the efficiency of prices is directly related to the cost of information, understanding the forces that make information costly is important. Our findings point out a new channel by which geography reduces information costs and thus broadens the role that geography plays in setting prices.

References

Almazan, Andres, Adolfo de Motta, and Sheridan Titman, 2010, Financial structure, acquisition opportunities, and firm locations, *Journal of Finance* 65, 529-563.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31-36.

Bayer, Patrick, Stephen L. Ross and Giorgio Topa, 2008, Place of work and place of residence: informal hiring network and labor market outcomes, *Journal of Political Economy* 116(6), 1150-1195.

Bhojraj, S., Charles M.C. Lee, and Derek Oler, 2003 What's My Line? A Comparison of Industry Classification Schemes for Capital Market Research, *Journal of Accounting Research* 41(5): 745-774.

Brennan, Michael J., Narasimhan Jegadeesh, Bhaskaran Swaminathan, 1993, Investment analysis and the adjustment of stock prices to common information, *Review of Financial Studies* 6, 799-824.

Brown, L., Hagerman, R., Griffin, P., and Zmijewski, M. 1987. Security Analysts Superiority Relative to Univariate Time-Series Models in Forecasting Quarterly Earnings. *Journal of Accounting and Economics*, 9(1): 61-87.

Brunnermeier, Marcus and Stefan Nagel. 2004. Hedge Funds and the Technology Bubble. *Journal of Finance,* 59(5), 2013-2040.

Bushee, Brian., and Theodore Goodman, 2007, Which Institutional Investors Trade Based on Private Information about Earnings and Returns?" *Journal of Accounting Research* 45, 1-31.

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

Christ, J 2009, New Economic Geography Reloaded: Localized Knowledge Spillovers and the Geography of Innovation, *Working Paper*, Universität Hohenheim.

Coval, Joshua, and Tobias Moskowitz, 1999, Home Bias at Home: Local Equity Preference in Domestic Portfolios, *Journal of Finance* 54, 2045-2073.

Coval, Joshua, and Tobias Moskowitz, 2001, The Geography of Investment: Informed Trading and Asset Prices, *Journal of Political Economy* 109, 811-841.

DeMarzo, Peter M., Dimitri Vayanos and Jeffrey Zwiebel, 2003, Persuasion bias, social influence, and unidimensional opinions, *Quarterly Journal of Economics* 118(3), 909-968.

Dougal, Casey, Christopher A. Parsons, and Sheridan Titman, 2012, Urban Vibrancy and Corporate Growth, *Working Paper*, University of Texas.

Ellison, Glenn., Edward Glaeser, and William Kerr, 2010, "What Causes Industry Agglomeration" Evidence from Coagglomeration Patterns. *American Economic Review,* forthcoming.

Engelberg, Joseph, Pengjie Gao and Christopher A. Parsons, 2012, The Price of a CEO's Rolodex, *Review Financial Studies* 26 (1): 79-114.

Engelberg, Joseph, Pengjie Gao and Ravi Jagannathan, 2008, An anatomy of pairs trading: the role of idiosyncratic news, common information and liquidity, Working paper.

Faccio, Mara and David C. Parsley, 2007, Sudden Deaths: Taking Stock of Geographic Ties, *Journal of Financial and Quantitative Analysis*, Forthcoming.

Grossman, Stanford and Joseph Stiglitz, 1980,On the impossibility of informationally efficient markets, *American Economic Review* 70, 393-408.

Grullon, Gustavo, George Kanatas and James Weston, 2004, Advertising, Breadth of Ownership, and Liquidity, *Review of Financial Studies* 17, 439-461.

Hameed, Allaudeen, Randall Morck, Jianfeng Shen and Bernard Yeung, 2008, Information markets, analysts, and comovement in stock returns, NYU Working paper.

Hong, Harrison, Jeffrey D. Kubik and Jeremy C. Stein, 2004, Social interaction and stock-market participation, *Journal of Finance* 59, 137-163.

Hou, Kewei, 2007, Industry information diffusion and the lead-lag effect in stock returns, *Review of Financial Studies* 20, 1113-1138.

Hou, Kewei and Tobias Moskowitz, 2005, Market frictions, price delay and the cross-section of expected stock returns, *Review of Financial Studies* 18, 981-1020.

Irvine, P. 2000. Do analysts generate trade for their firms? Evidence from the Toronto stock exchange. *Journal of Accounting and Economics* 30: 209-226.

Jacobs, J. 1969, *The Economy of Cities*, Random House, New York, Vintage.

Jegadeesh, N., Kim, J., Krische, S.D., and Lee, C.M.C. 2004. Analyzing the analysts: When do recommendations add value? *Journal of Finance,* 59: 1083-1124.

Kedia, Simi and Shivaram Rajgopal, 2008, Neighborhood matters: The impact of location on broad based stock options plans, *Journal of Financial Economics*, forthcoming.

Khan, Mozaffar and Hai Lu. 2011, Do short sellers front-run insider sales? *Working Paper, Massachusetts Institute of Technology*.

Lakonishok, Josef, Andrei Shleifer, Richard Thaler, and Robert Vishny, 1991, Window Dressing by Pension Fund Managers, *American Economic Review 81*, 227-231.

Lee, Charles and Bhaskaran Swaminathan, 2000, Price Momentum and Trading Volume, *Journal of Finance 55*, 2017-2070.

Lo, Andrew W. and A. Craig MacKinlay, 1990, When are contrarian profits due to stock market overreaction? *Review of Financial Studies* 3, 175-205.

Loughran, Tim and Paul Schultz, 2005, Liquidity: Urban versus rural firms, *Journal of Financial Economics* 78, 341-374.

Mech, Timothy S., 1993, Portfolio return autocorrelation, *Journal of Financial Economics* 34, 307-344.

Pirinsky, Christo and Qinghai Wang, 2006, Does corporate headquarters location matter for stock returns?, Journal of Finance 61, 1991-2015.

Rhodes-Kropf, Matthew, David Robinson, and S. Viswanathan, 2004, Market valuation and merger waves, *Journal of Finance* 59, 2685-2718.

Scharfstein, David and Jeremy Stein 1990, Herd behavior and investment, *American Economic Review* 80, 465-479.

Van Nieuwerburgh, Stijn and L. Veldkamp 2010, Information Acquisition and Under-Diversification, *Review of Economic Studies* 77(2), 779-805.

Veldkamp, Laura, 2006, Information markets and the comovement of asset prices, *Review of Economic Studies* 73, 823-845.

Welch, Ivo, 1992, Sequential sales, herding and cascades, *Journal of Finance* 47, 695-732.

Yan, Xuemin and Zhe Zhang, 2007. Institutional Investors and Equity Returns: Are Short-term Institutions Better Informed? *The Review of Financial Studies*, 22, 893-924.

Table 1: Summary Statistics

Table 1 provides the descriptive statistics for the sample firms. 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 CRSP 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 previous year's total assets. Sales growth is the annual percentage change in sales. Tangible assets are net fixed assets, defined as plant, property and equipment. Capital expenditure, capital expenditure/R&D, tangible assets and cash are scaled by 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 measure is the average daily absolute return over daily trading volume, constructed as in Amihud (2002).

	Mean	Median	Standard Deviation	10th Percentile	90th Percentile
Cluster ratio	0.10	0.07	0.09	0.01	0.22
Assets	341.77	87.74	623.53	11.83	992.58
Market capitalization	1723.56	109.47	11434.36	9.29	1965.78
Sales	338.07	78.49	634.46	6.931	1005.83
Capital Expenditure	0.08	0.04	0.15	0.01	0.16
Capital Expenditure + R&D	0.19	0.13	0.22	0.04	0.39
ROA	0.07	0.11	0.27	-0.20	0.29
Sales Growth	0.25	0.10	1.38	-0.20	0.65
SGA/Sales	0.49	0.31	1.01	0.10	0.85
Tangible Assets	0.26	0.17	0.32	0.04	0.61
Market-to-book	2.62	1.68	4.34	0.91	4.87
R&D	0.13	0.07	0.2	0	0.31
Adv. Exp/ Sales	0.06	0.021	0.36	0.00	0.1
Cash	0.29	0.15	0.46	0.02	0.68
Number of Employees	6.08	0.45	29.81	0.044	9.7
Trading Volume	4184.13	390.31	20517.87	14.97	7074.40
Illiquidity	0.40	0.03	1.12	0.00	1.15
Turnover	1.58	1.06	1.73	0.26	3.50

Table 2: Cluster/Out-of-Cluster Differences

Table 2 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 ten or more firms with the same 3-digit SIC code located in the same 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 CRSP 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 previous year's total assets. Sales growth is the annual percentage change in sales. Tangible assets are net fixed assets, defined as plant, property and equipment. Capital expenditure, capital expenditure/R&D, tangible assets and cash are scaled by 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 measure is the average daily absolute return over daily trading volume, constructed as in Amihud (2002). The t-statistics are clustered by firm. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

	Mean: All Firms	Mean: Firms Outside Clusters	Mean: 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	1723.56	1490.66	2175.61	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	4184.13	2656.85	7881.35	5224.50***	5.47
Illiquidity	0.40	0.49	0.19	-0.30***	-12.4
Turnover	1.58	1.31	2.25	0.94***	17.72

Table 3: C0-Movement in Fundamentals

ΔEarnings is the change in firm profitability, measured as the annual change in earnings before interest, taxes and amortization scaled by previous year's total assets. *ΔSales* is the annual change in a firm's sales, scaled by previous year's sales. $\Delta(Capx + R\&D)$ is the annual change in a firm's total capital expenditure, including R&D spending, scaled by previous year's total assets. *ΔR&D* is the annual change in a firm's R&D spending, scaled by previous year's total assets. *ΔSGA* is the annual change in a firm's SGA Expenses, scaled by the previous year's total assets. $\Delta Cash$ is the annual change in a firm's total cash holdings, scaled by previous year's total assets. $\Delta Earnings_{MSA-IND}$ is the equally-weighted change in profitability for same-industry firms within the MSA, and $\Delta Earnings_{IND}$is the equally-weighted change in profitability for the industry, as defined by 3-digit SIC code. All other MSA-wide and industry-wide average changes in fundamentals are defined analogously. Cluster is a dummy variable that takes a value of one if the firm is located in an industry cluster. Standard errors are clustered by firm. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

Dependent variable:	Δ(Earnings)		Δ(Sales)		Δ(CAPX+R&D)		Δ(R&D)		Δ(SGA)		Δ(CASH)	
$\Delta Fundamental_{MSA-IND}$	0.030*	-0.001	0.028*	-0.005	0.079***	0.020	0.066**	0.009	0.057***	-0.010	-0.032	-0.063***
	(0.018)	(0.018)	(0.015)	(0.014)	(0.023)	(0.029)	(0.027)	(0.028)	(0.020)	(0.017)	(0.022)	(0.021)
$\Delta Fundamental_{MSA-IND}$ x Cluster		0.104**		0.108**		0.138***		0.119**		0.208***		0.120**
		(0.047)		(0.043)		(0.050)		(0.053)		(0.059)		(0.049)
$\Delta Fundamental_{IND}$	0.531***	0.366***	0.182***	0.077	0.219**	0.201**	0.387***	0.286***	0.431***	0.227***	0.739***	0.544***
	(0.046)	(0.058)	(0.055)	(0.079)	(0.098)	(0.082)	(0.083)	(0.107)	(0.063)	(0.057)	(0.080)	(0.127)
$\Delta Fundamental_{IND}$ x Cluster		0.261***		0.130		-0.030		0.105		0.285***		0.221
		(0.091)		(0.093)		(0.105)		(0.128)		(0.092)		(0.151)
Cluster		-0.004		-0.055		0.000		-0.003		-0.019***		-0.019**
		(0.004)		(0.036)		(0.004)		(0.003)		(0.005)		(0.009)
R-square	0.020	0.022	0.014	0.022	0.033	0.034	0.022	0.023	0.067	0.073	0.035	0.033
# of observations	22244	22244	24550	24550	16694	16694	18105	18105	20220	20220	22714	22714
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Table 4: Analysis of Analyst Coverage and Industry Clusters

Analyst Coverage is measured as the aggregate number of unique year-end estimates found within the IBES Detail History file. In Panel A, Cluster ratios of low, medium, and high represent firms who 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 who is headquartered in an MSA with at least 10 other firms within the same 3-digit 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 one when a firm is headquartered within an MSA with at least 10 other firms within the same 3-digit SIC code. *Size* is the market value of equity computed from CRSP 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. *Pin* values are calculated as the yearly average of the quarterly PIN values provided by Stephen Brown according to the methodology derived by Easley, Kiefer, and O'Hara (1997). *Ret* is calculated as the 6-month returns prior to the end of the fiscal year. LogAdv is the sum of a firm's annual advertising and SG&A expenses, scaled by annual sales. *BEME* is the ratio of book equity to market value of equity. Standard errors are clustered by firm. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

PANEL A: Univariate Analysis

	Dependent Variable = Analyst Following		
	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***		11.87
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 (1 + Analyst Following)	
Cluster Ratio	0.140**	
	(0.058)	
Cluster Dummy		0.054***
		(0.009)
LogSize	0.447***	0.448***
	(0.013)	(0.013)
Prc	-0.002***	-0.002***
	(0.000)	(0.000)
Pin	-1.110***	-1.104***
	(0.267)	(0.267)
Ret	-0.610***	-0.611***
	(0.144)	(0.144)
LogAdv	-0.009	-0.009
	(0.015)	(0.015)
LogBEME	0.100***	0.099***
	(0.010)	(0.010)
Year fixed effects	Yes	Yes
Industry Fixed Effects	Yes	Yes
MSA Fixed Effects	Yes	Yes
Number of Observations	21171	21171
R-square	0.7091	0.7093

Table 5: Analysis of Institutional Holdings and Industry Clusters

BET equals one if the percent of the institution's equity portfolio invested in a firm is in the top quintile of its total holdings for the given year and zero otherwise. In Panel A, Cluster ratios of low, medium, and high represent firms who 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 who is headquartered in an MSA with at least 10 other firms with the same 3-digit 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 one when a firm is headquartered within an MSA with at least 10 other firms within the same 3-digit SIC code. *LogSize* is the logarithm of firm size, measured as market capitalization. *LogNumEst* is the log of one 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 IBES Detail History file. *Stdev* is the standard deviation of analysts' earnings forecasts for the upcoming year. *LogVol* is the 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. *LogIlliq* is the logarithm of the average daily absolute return over daily trading volume, constructed as in Amihud (2002). T-statistics in logistic regressions are clustered by firm and year. *, **, and *** represent 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		
	Mean	Standard Deviation	t-stat
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	0.4038	
Inside Cluster	0.2270	0.41891	
Inside – Outside	0.02196***		40.83

Panel B: Logistic Regression Analyses

	Dependent Variable = BET	
Cluster Ratio	0.181***	
	(0.04400)	
	1.19800	
Cluster Dummy		0.0385***
		(0.01230)
		1.03900
LogSize	0.5164***	0.5161***
	(0.00741)	(0.00740)
	1.67600	1.67500
LogNumEst	0.0319**	0.0329***
	(0.00992)	(0.00997)
	1.03200	1.03300
Stdev	-0.0088	-0.00883
	(0.00594)	(0.00598)
	0.99100	0.99100
LogVOL	-0.0960***	-0.0941***
	(0.01140)	(0.01130)
	0.90800	0.91000
PRC	0.00124***	0.00123
	(0.00027)	(0.00027)
	1.00100	1.00100
Pin	-0.2059***	-0.2004**
	(0.07860)	(0.13990)
	0.81400	0.81800
Ret	0.6977***	0.6976***
	(0.04240)	(0.04240)
	2.00900	2.00900
logIlliq	-0.0490	-0.0481
	(0.01260)	(0.01250)
	0.95200	0.95300
Year fixed effects	Yes	Yes
Industry Fixed Effects	Yes	Yes
Number of Observations	2421900	2421900

Table 6: Institutional Holdings and Small Firms in Industry Clusters

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 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 *** represent significance at the 10%, 5% and 1% levels, respectively.

	Dependent Variable: Number of Small Cluster Holdings		
Num_Large_BET_Inside	1.0931***	0.7713**	0.8496**
	(0.3214)	(0.3088)	(0.3200)
Num_Large_BET_Outside	0.1138***	0.1039***	0.1066***
	(0.0152)	(0.0067)	(0.0066)
Number of Clusters	61	61	61
Industry Fixed Effects	Yes	Yes	Yes
Year fixed effects	Yes	Yes	Yes
Number of Observations	1676352	79638	71089
R-square	0.231	0.083	0.092
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

Table 7: Industry Clusters and Price Delay

IND1 is the industry-wide price delay measure constructed as in equation (4) in text. *Cluster Ratio* is defined as the number of firms in a given industry, as defined by the 3-digit SIC code, and in the same MSA code, scaled by the total number of firms in the same industry. *LogSize* is the logarithm of firm size, measured as market capitalization. *LogNumEst* is the log of one 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 IBES Detail History file. *LogNumInst* is the average of the total number of institutions currently holding the stock over the calendar year. *LogVol* is the logarithm of the yearly average of monthly dollar trading volume. *LogTurnover* is the logarithm of the yearly average of the monthly number of shares traded divided by the number of shares outstanding. *LogIlliq* is the logarithm of the average daily absolute return over daily trading volume, constructed as in Amihud (2002). Standard errors are clustered by industry. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

	Dependent Variable: IND1						
Cluster Ratio	-0.212***	-0.161***	-0.059***	-0.047***	-0.043***	-0.029**	-0.038***
	(0.0229)	(0.0264)	(0.0130)	(0.0097)	(0.0096)	(0.0105)	(0.0206)
LogSize			-0.041***	-0.025***	-0.007***	0.004*	0.004*
			(0.0013)	(0.0013)	(0.0016)	(0.0021)	(0.0021)
LogNumEst				-0.042***	-0.025***	-0.019***	-0.019***
				(0.0028)	(0.0028)	(0.0027)	(0.0027)
LogNumInst					-0.038***	-0.032***	-0.032***
					(0.0017)	(0.0018)	(0.0019)
LogVol						-0.007**	-0.008***
						(0.0032)	(0.0033)
LogTurnover						0.005	0.005
						(0.0032)	(0.0033)
LogIlliq						0.010***	0.010***
						(0.0013)	(0.0013)
Year Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
Industry Fixed Effects	No	Yes	Yes	Yes	Yes	Yes	Yes
MSA Fixed Effects	No	No	No	No	No	No	Yes
Number of Observations	49759	49759	49345	49345	48240	48237	48237
R-square	0.011	0.062	0.264	0.279	0.287	0.292	0.297

Table 8: Analysis of Comovement in Fundamentals for Relocating Firms

ΔEarnings is the change in firm profitability, measured as the annual change in earnings before interest, taxes and amortization scaled by previous year's total assets. *ΔSales* is the annual change in a firm's sales, scaled by previous year's sales. $\Delta(Capx + R\&D)$ is the annual change in a firm's total capital expenditure, including R&D spending, scaled by previous year's total assets. *ΔR&D* is the annual change in a firm's R&D spending, scaled by previous year's total assets. *ΔSGA* is the annual change in a firm's SGA Expenses, scaled by the previous year's total assets. $\Delta Cash$ is the annual change in a firm's total cash holdings, scaled by previous year's total assets. Post-move represents the two calendar years following the year of the move, while pre-move represents the two calendar years prior to the year of the move. Move year represents the calendar year of the firm's move. Pearson correlation coefficients are tabulated in Panel A. In Panel B, $\Delta Fundamental_{MSA,IND,OLD}$ is the equally-weighted change in the firm's fundamentals for same-industry firms within the MSA prior to the move, and $\Delta Fundamental_{MSA,IND,NEW}$ is the equally-weighted change in the firm's fundamentals post-move. $\Delta Fundamental_{IND}$ is the equally-weighted change in the industry's fundamentals as classified by SIC3 code. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

PANEL A: Correlation of Fundamentals with Post-Move Industry Cluster

	Pre-Move (t-1, t-2)	Move Year	Post-Move (t+1, t+2)
ΔCAPX	0.182**	0.238**	0.351***
Δ(CAPX + R&D)	0.05888	0.191*	0.277***
ΔR&D	-0.1389	-0.04158	0.10321
ΔSG&A	0.11045	-0.09287	0.273***
ΔEarnings	0.0625	0.14966	-0.07037
ΔSales	0.19601	0.01454	0.0892
ΔCash	0.213**	0.15238	0.00991

PANEL B: Regression Analysis of Comovement in Fundamentals for Relocating Firms

Dependent variable	Δ(CAPX)	Δ(CAPX + R&D)	Δ(R&D)	Δ(SGA)	ΔEarnings	ΔSales	ΔCash
$\Delta Fundamental_{MSA,IND,OLD}$	0.005	0.016	-0.386	0.023	0.042	-0.521	-0.668
	(0.135)	(0.128)	(0.305)	(0.168)	(0.275)	(0.313)	(0.246)
Postmove	-0.012	-0.022	-0.029	-0.031	0.010	0.023	0.019
	(0.007)	(0.014)	(0.020)	(0.022)	(0.028)	(0.119)	(0.044)
$\Delta Fundamental_{MSA,IND,OLD}$ x PostMove	0.053	0.053	0.439	-0.146	-0.078	0.443	0.504*
	(0.155)	(0.162)	(0.379)	(0.245)	(0.316)	(0.350)	(0.293)
$\Delta Fundamental_{MSA,IND,NEW}$	-0.204	-0.617	-0.895	-0.240	-0.227	0.571	0.088
	(0.149)	(0.332)	(0.465)	(0.278)	(0.306)	(0.449)	(0.272)
$\Delta Fundamental_{MSA,IND,NEW}$ x PostMove	0.353*	0.893**	1.419**	0.080	0.032	-0.623	-0.246
	(0.206)	(0.391)	(0.578)	(0.263)	(0.326)	(0.459)	(0.364)
$\Delta Fundamental_{IND}$	1.083***	1.148**	-0.908	0.062	1.081***	0.530	1.499***
	(0.339)	(0.476)	(0.964)	(0.437)	(0.357)	(0.472)	(0.481)
Number of Observations	277	277	158	256	279	274	279
Adjusted R-square	0.167	0.094	0.100	0.164	0.046	0.125	0.084
Year fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Industry fixed effects	Yes	Yes	Yes	Yes	Yes	Yes	Yes
MSA fixed effects	No	No	No	No	No	No	No

Table 9: Institutional Holdings and Analyst Coverage for Relocating Firms

ChgBET calculates the change in the aggregate number of BET institutions before and after the firm's location. *ChgBETPCT* calculates the percent change in the number of BET institutions as scaled by the BET prior to the firm's move before and after the firm's relocation. *BET* equals one if the percent of the institution's equity portfolio invested in a firm is in the top quintile of its total holdings for the given year and zero otherwise. *ChgLogNumEst* is the change in amount of analyst coverage before and after the firm's relocation, measured as the aggregate number of year-ahead earnings estimates as found in the IBES Detail History file. *Relocation Dummy* takes a value of one if the firm moves into a larger cluster, as measured by Cluster Ratio, and a value of minus one if the firm moves into a smaller cluster. Standard errors are clustered by firm. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

PANEL A: Univariate Analysis

	ChgBET		ChgBETPCT		ChgLogNumEst	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Firm moves into smaller clusters	2.11	(3.14)	2.3%	(0.21)	0.18	(0.66)
Firm moves into larger clusters	7.24	(2.95)	77.4%	(0.38)	0.26	(0.54)
Difference	5.13		75.05%**		0.08	

Panel B: Regression analysis

Dependent variable	ChgBET		ChgBETPCT		ChgLogNumEst	
Intercept	4.51**		0.552**		0.217	
	(1.93)		(0.20)		(0.16)	
Relocation Dummy	2.61	4.80**	0.328*	0.424**	0.042	0.153**
	(2.15)	(2.05)	(0.17)	(0.20)	(0.156)	(0.06)
Number of Observations	59	59	59	59	21	21
Adjusted R-Squared	0.020	0.400	0.046	0.282	0.040	0.158
Year Fixed Effects	No	Yes	No	Yes	No	Yes

Table 10: Price Delay for Relocating Firms

IND1, IND2 and *IND3* are the industry-wide price delay measure constructed as in equations (4), (5) and (6) 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 one in the calendar year following the year of the relocation. Standard errors are clustered by firm. *, **, and *** represent significance at the 10%, 5% and 1% levels, respectively.

	Moves into smaller clusters			Moves into larger clusters		
Dependent 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)
Number of Observations	71	71	71	117	117	117
Adjusted R-Squared	0.213	0.200	0.199	0.225	0.245	0.243
Year Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes

Table 11: Tests of Industry Clusters as Finer Industry Codes

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

3-Digit SIC Code	Increased Probability of Same SIC4 within Cluster	Standard Error	T-Stat		3-Digit SIC Code	Increased Probability of Same SIC4 within Cluster	Standard Error	T-Stat	
131	-0.00142	0.00527	-0.2694		381	-0.0509	0.0481	-1.0582	
138	-0.0198	0.00859	-2.3050	**	382	0.00611	0.01	0.6110	
201	-0.1302	0.1144	-1.1381		384	-0.041	0.00513	-7.9922	***
208	0.07	0.0516	1.3566		394	-0.129	0.0507	-2.5444	***
209	0.0672	0.055	1.2218		399	-0.0689	0.0383	-1.7990	*
251	-0.1282	0.1371	-0.9351		504	-0.1153	0.0192	-6.0052	***
271	-0.00322	0.0789	-0.0408		506	0.0462	0.0269	1.7175	*
281	0.1456	0.0576	2.5278	***	508	-0.0104	0.0377	-0.2759	
282	0.0596	0.0947	0.6294		509	0.0123	0.0344	0.3576	
283	0.0124	0.00333	3.7237	***	512	0.1652	0.0432	3.8241	***
284	0.0775	0.0264	2.9356	***	513	0.0958	0.0264	3.6288	***
289	0.0386	0.0491	0.7862		514	-0.0747	0.0643	-1.1617	
291	0.0438	0.041	1.0683		541	-0.0305	0.0731	-0.4172	
314	0.0994	0.0825	1.2048		581	0.0387	0.0146	2.6507	***
331	-0.0324	0.0395	-0.8203		596	-0.0178	0.0264	-0.6742	
344	0.0508	0.0744	0.6828		599	-0.0427	0.0308	-1.3864	
349	0.0582	0.0505	1.1525		701	0.0633	0.0265	2.3887	***
353	0.2023	0.0262	7.7214	***	731	0.0298	0.0238	1.2521	
355	0.0495	0.0234	2.1154	**	735	-0.1209	0.0514	-2.3521	***
356	-0.0255	0.0282	-0.9043		736	-0.0148	0.0238	-0.6218	
357	0.0519	0.00488	10.6352	***	737	0.0164	0.00158	10.3797	***
358	0.00811	0.0577	0.1406		738	0.00613	0.00915	0.6699	
362	-0.0844	0.0541	-1.5601		799	-0.038	0.0238	-1.5966	
364	0.0233	0.0593	0.3929		806	0.0863	0.0494	1.7470	*
365	0.0256	0.0387	0.6615		807	-0.0219	0.0577	-0.3795	
366	-0.00796	0.00552	-1.4420		808	0.1441	0.0376	3.8324	***
367	0.0899	0.00377	23.8462	***	809	0.1055	0.0168	6.2798	***
369	0.0562	0.0353	1.5921		873	-0.0318	0.0112	-2.8393	***
371	-0.0366	0.0222	-1.6486	*	874	0.0474	0.0223	2.1256	**
372	-0.0475	0.0421	-1.1283						
					Mean	**0.0130**	**0.0386**	**0.3372**	